

**UNIVERSIDADE FEDERAL DE PELOTAS**  
Centro de Desenvolvimento Tecnológico  
Programa de Pós-Graduação em Computação



Dissertação

***Early Skip/DIS: Uma Heurística para Redução de Complexidade no Codificador de Mapas de Profundidade do 3D-HEVC***

**Ruhan Avila da Conceição**

Pelotas, 2016

**Ruhan Avila da Conceição**

***Early Skip/DIS: Uma Heurística para Redução de Complexidade no Codificador de Mapas de Profundidade do 3D-HEVC***

Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação

Orientador: Prof. Dr. Luciano Volcan Agostini  
Coorientadores: Prof. Dr. Bruno Zatt  
Prof. Dr. Marcelo Schiavon Porto

Pelotas, 2016

**Insira AQUI a ficha catalográfica  
(solicitada na página da biblioteca)**

**Insira AQUI a folha de aprovação  
(será entregue após a banca)**

**Dedico este trabalho à minha esposa Bruna.**

## **AGRADECIMENTOS**

Ao escrever esta dissertação de mestrado, sinto-me imensamente alegre por estar concludendo mais uma etapa de minha vida. Durante este ano de curso, tive a oportunidade de conhecer e conviver com pessoas que foram essenciais em minha jornada acadêmica, possibilitando a concretização deste trabalho.

Agradeço a Deus por me proporcionar um crescimento profissional e pessoal ao longo deste ano, colocando-me no rumo das pessoas certas; por livrar-me do mal e das enfermidades; e por jamais deixar faltar-me o sustento.

À minha esposa Bruna. O amor, carinho, companheirismo e afeto que compartilhamos me auxiliaram durante este ano de mestrado, e com certeza, serão essenciais no enfrentamento dos desafios que sobrevirão futuramente em minha vida pessoal, acadêmica e profissional. Eu te amo.

Aos meus pais, Altamir e Rosane Conceição, pelo amor e dedicação que tiveram para comigo; pelo apoio, carinho e conselhos que me foram essenciais ao longo de toda a minha vida. Agradeço pelas lições inestimáveis de ética e moral que me deram, não apenas através de conselhos, mas também de atitudes. Estas foram fundamentais na formação de meu caráter.

Ao meu irmão Josuan Conceição, por ser-me um exemplo de aluno e profissional através do seu esforço e responsabilidade. Sou grato por ser o meu primeiro orientador na vida acadêmica, o qual me auxilia desde quando eu cursava as séries iniciais da escola.

Aos meus avós Jorge (*in memoriam*) e Helena Conceição, Tailor (*in memoriam*) e Nilza Avila, sou grato por estarem presentes, sempre que possível, em todos os momentos de minha vida. Desejo tornar-me um profissional que honre os valores de ética e moral que vocês ensinaram aos meus pais, e estes a mim.

A todos os meus demais familiares que de alguma maneira contribuíram para a minha formação. Em especial, sou grato aos meus tios Raquel e Roberto, Jorginho, Honeide e Magna e, Sidnei e Rose, pelo carinho, apoio, conselhos, momentos de descontração etc. que marcaram a vida em todos os aspectos.

Aos meus amigos, em especial ao Érico, Adriano, Jonathas, Filipe, Matheus, Rafael, Francescoli por diversas vezes me aconselharem e ouvirem minhas angústias. Agradeço, pelos momentos de lazer que desfrutamos, como os campeonatos de vídeo game, os chimarrões, as noites viradas na internet, etc. Enfim, agradeço pelas lembranças que jamais serão esquecidas.

Ao meu orientador, o professor Luciano Agostini, pelos direcionamentos, conselhos e advertências que recebi desde quando fui bolsista do grupo PET computação e de iniciação científica no GACI, os quais foram fundamentais no meu crescimento pessoal e acadêmico. Também sou grato aos professores Marcelo Porto e Bruno Zatt, que,

na figura de meus coorientadores, foram fundamentais em minha jornada científica e acadêmica ao longo deste ano de mestrado.

Às pessoas com quem compartilhei boa parte de meu tempo durante minha graduação e mestrado. Pessoas estas que foram mais do que simplesmente colegas; foram, e ainda são, meus amigos. Em especial, agradeço ao Cláudio Júnior, Gustavo Wrege, Henrique Maich, Guilherme Paim, Jones Göebel, Vladimir Afonso e Guilherme Corrêa pela amizade desfrutada, pelos momentos de descontração no laboratório de pesquisa e pelos trabalhos realizados e publicados em conjunto. Além destes, sou imensamente grato ao Giovanni Avila por auxiliar-me no processo de simulação e re-colhimento de resultados das soluções propostas nesta dissertação.

Por fim, a todos que, de uma maneira ou de outra, me auxiliaram para que esta dissertação de mestrado fosse concluída e que, por um relapso de minha mente, não foram citadas acima.

**Combattei o bom combate, acabei a carreira, guardei a fé.**  
— II TIMÓTEO 4:7

## RESUMO

CONCEIÇÃO, Ruhan Avila da. ***Early Skip/DIS: Uma Heurística para Redução de Complexidade no Codificador de Mapas de Profundidade do 3D-HEVC.*** 2016. 95 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2016.

Vídeos 3D provêm uma experiência visual elevada aos espectadores devido à percepção de profundidade das imagens. Apesar disto, o tamanho destes vídeos tende a crescer linearmente conforme o número de vistas codificadas, considerando formatos de vídeos convencionais. Neste cenário surge o formato *Multiview plus Depth* (MVD), o qual associa informações de distância entre os objetos da cena e a câmera (mapas de profundidade), permitindo um processo eficiente de síntese de vistas intermediárias, reduzindo o número de vistas a serem transmitidas. Ao contrário de padrões multivistas anteriores, o 3D-HEVC é capaz de processar mapas de profundidade, criando novas ferramentas para manipulá-los e codificá-los. Embora este fato proporcione um aumento na eficiência de compressão, o acréscimo de novas ferramentas no codificador acarreta no aumento da complexidade do processo. Assim, cresce a relevância de soluções que reduzam o tempo de codificação do 3D-HEVC, sem impactar significativamente a eficiência de codificação. Este trabalho apresenta uma heurística de redução de complexidade para o codificador de mapas de profundidade do 3D-HEVC, chamada de *Early Skip/DIS*. Uma análise sobre mapas de profundidade do 3D-HEVC é apresentada nesta dissertação, demonstrando que o particionamento  $2Nx2N$  é largamente utilizado pelo codificador, visto que diversas ferramentas eficientes de codificação, utilizam exclusivamente este modo. A análise demonstrou que, além do  $2Nx2N$  ser o modo de particionamento mais usado, a exclusão dos demais modos gera um impacto desprezível em eficiência de codificação, com ganhos mínimos em termos de tempo de processamento. Este fato conduziu ao desenvolvimento da heurística *Early Skip/DIS*, a qual visa evitar o teste dos demais modos/ferramentas de predição com base no custo RD gerado por estes modos. Os *thresholds* utilizados nesta solução são definidos de forma adaptativa. Resultados de simulação demonstraram que a solução é capaz de reduzir o tempo de codificação dos mapas de profundidade em até 33,7%, com um impacto médio de apenas 0,047% na eficiência de compressão da textura. A heurística proposta apresenta os melhores resultados de redução de complexidade para o codificador de mapas de profundidade entre os trabalhos relacionados.

**Palavras-chave:** Vídeos 3D, 3D-HEVC, Mapas de Profundidade, Redução de Complexidade.

## ABSTRACT

CONCEIÇÃO, Ruhan Avila da. **Early Skip/DIS: A Complexity-Reduction Heuristic for 3D-HEVC Depth Coder.** 2016. 95 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2016.

3D videos provides a visual experience with depth perception through the usage of special displays that project a three-dimensional scene from slightly different directions for the left and right eyes. Despite this improved visual experience, the coded-video data volume tends to linearly increases with the number of processed views, mainly considering conventional 3D video formats. In this scenario emerges the Multiview plus Depth (MVD) format, which informs the distance between scene objects and the recording camera (depth maps), allowing an efficiently view-synthesis process while reducing the number of views to be transmitted. Unlike previous multiview video coding standards, 3D-HEVC is able to manipulate depth maps in an efficient way due the new defined tools which explores the depth maps characteristics. Although this fact leads to an improvement of 3D-HEVC compression efficiency, the addition of new coding tools also increases the coding process complexity. Thus, solutions, which reduces the 3D-HEVC coding time while does not affecting the compression efficiency at all, are important in this scenario. This work presents a complexity reduction heuristic for the 3D-HEVC depth maps coder, called early Skip/DIS. Initially, an analysis about 3D-HEVC depth-maps coder is presented. This analysis showed that the 2Nx2N is the most used partitioning mode, since some efficient coding tools, like Skip and DIS, are applied exclusively over this partitioning mode. This analysis also showed that, beyond the 2Nx2N partitioning mode is the most used mode, the exclusion of the other partition modes causes an imperceptible impact in the encoding efficiency and a low impact in processing time. This fact leads to the development of an early decision heuristic called early Skip/DIS, which avoids the encoder checking unnecessary modes based on the RD cost generated by the Skip and DIS modes. The thresholds used in this solution are defined in an adaptively way, observing the occurrence rate of those modes as a function of its generated RD costs. Simulation results demonstrated that the proposed solution is able to reduce the depth-map coding time up to 33.7% while affecting the texture compression efficiency in 0.047% (in terms of BD-rate). The propose heuristic presented the best depth-map complexity reduction result among other related works.

**Keywords:** 3D Videos, 3D-HEVC, Depth Maps, Complexity Reduction.

## LISTA DE FIGURAS

Figura 1	Diagrama de blocos de alto nível para o codificador de vídeos 3D-HEVC . . . . .	21
Figura 2	Diagrama de blocos do codificador de vídeos HEVC. Adaptada de (SULLIVAN et al., 2012) . . . . .	27
Figura 3	Particionamento de CUs em PUs . . . . .	28
Figura 4	Subdivisão de uma CTU em CUs (linhas contínuas) juntamente subdivisão de CUs folhas em TUs (linhas pontinhadas). (a) CTU e suas subdivisões internas. (b) <i>Quadtree + RQT</i> correspondente . . . . .	29
Figura 5	33 modos direcionais e seus respectivos ângulos da predição intraquadro (ROSEWARNE et al., 2015) . . . . .	30
Figura 6	Mapeamento entre as direções da predição intraquadro e os modos de predição (ROSEWARNE et al., 2015) . . . . .	30
Figura 7	Exemplo de aplicação do modo 26 de predição intraquadro. (a) bloco original (b) bloco predito . . . . .	31
Figura 8	Funcionamento da predição interquadros do HEVC. Adaptada de (SZE; BUDAGAVI; SULLIVAN, 2014) . . . . .	32
Figura 9	Posições de candidatos espaciais do <i>Merge</i> (ROSEWARNE et al., 2015) . . . . .	33
Figura 10	Possíveis posições para o candidato temporal do <i>Merge</i> em relação ao bloco co-localizado Y (ROSEWARNE et al., 2015) . . . . .	33
Figura 11	Exemplo de um sistema com duas vistas e seus respectivos mapas de profundidade . . . . .	38
Figura 12	Visão geral da estrutura e formato de dados para transmissão de vídeos 3D. Adaptada de (CHEN et al., 2015) . . . . .	39
Figura 13	Estrutura de Unidades de Acesso do 3D-HEVC. Adaptada de (CHEN et al., 2015) . . . . .	40
Figura 14	Estrutura básica do codificador 3D-HEVC com informações inter-vista e inter-componente. Adaptada de (CHEN et al., 2015) . . . . .	41
Figura 15	Localização dos blocos vizinhos temporais e espaciais do NBDV. Adaptada de (CHEN et al., 2015) . . . . .	42
Figura 16	Processo de obtenção dos parâmetros de movimento para um bloco no quadro atual baseado em parâmetros de movimento da vista de referência já codificada e uma estimativa do mapa de profundidade para o quadro atual. Adaptada de (CHEN et al., 2015) . . . . .	44

Figura 17	DMM <i>Wedgelet</i> de um bloco de profundidade: (a) espaço contínuo, (b) espaço discreto, (c) padrão de partição correspondente (CHEN et al., 2015) . . . . .	47
Figura 18	DMM <i>Countour</i> de um bloco de profundidade: (a) espaço contínuo, (b) espaço discreto, (c) padrão de partição correspondente (CHEN et al., 2015) . . . . .	48
Figura 19	Particionamento da textura e os possíveis particionamentos de profundidade. Adaptada de (CHEN et al., 2015) . . . . .	50
Figura 20	Modos de predição intraquadro (a) vertical e (b) horizontal utilizados pela ferramenta DIS . . . . .	50
Figura 21	Amostras vizinhas espaciais reconstruídas utilizadas pela ferramenta DIS . . . . .	51
Figura 22	Modos de particionamento de CUs em PUs classificadas de acordo com seu formato . . . . .	52
Figura 23	Distribuição dos modos de particionamento de PUs de mapas de profundidade para diversas sequências de teste utilizando (a) QP 34 e (b) QP 45 . . . . .	54
Figura 24	Distribuição em escala logarítmica dos modos de particionamento de PUs de mapas de profundidade para diversas sequências de teste utilizando (a) QP 34 e (b) QP 45 . . . . .	54
Figura 25	Distribuição dos modos de particionamento de PUs classificados em retangulares e assimétricos de mapas de profundidade para diversas sequências de teste utilizando (a) QP 34 e (b) QP 45 . . . . .	55
Figura 26	Taxa de ocorrência das ferramentas de predição <i>Skip</i> , <i>Merge</i> Apenas e DIS, utilizando QP (a) 34 e (b) 45 . . . . .	56
Figura 27	Taxa de ocorrência dos modos (a) <i>Skip</i> e (b) DIS em função do custo RD gerado . . . . .	66
Figura 28	Análise da taxa de ocorrência dos modos (a) <i>Skip</i> e (b) DIS da sequência de vídeo <i>Kendo</i> e dos modos (c) <i>Skip</i> e (d) DIS da sequência <i>Undo_Dancer</i> . . . . .	67
Figura 29	Exemplo de definição de <i>thresholds</i> para heurística <i>Early Skip/DIS</i> utilizando (a) THR 95% e (b) THR 90% . . . . .	71
Figura 30	Fluxograma de obtenção dos <i>Thresholds</i> . . . . .	71
Figura 31	Fluxograma da heurística <i>early Skip/DIS</i> . . . . .	72
Figura 32	Gráfico comparativo entre resultados de redução de tempo (a) e perda na eficiência de codificação (b) obtidos através da heurística <i>Early Skip/DIS</i> utilizando THR 99%, 95%, 90% e 80%. . . . .	79

## **LISTA DE TABELAS**

Tabela 1	Resultados de simulação da análise RDC para o SA-HTM . . . . .	59
Tabela 2	Resultados de simulação da análise RDC para o SR-HTM . . . . .	60
Tabela 3	Comparação entre os resultados médios da análise RDC de SA-HTM e SR-HTM . . . . .	61
Tabela 4	Número total de quadros, taxa de exibição e tempo total das sequências de teste . . . . .	73
Tabela 5	Resultados de simulação da heurística <i>early Skip/DIS</i> utilizando THR 99% . . . . .	74
Tabela 6	Resultados de simulação da heurística <i>early Skip/DIS</i> utilizando THR 95% . . . . .	75
Tabela 7	Resultados de simulação da heurística <i>Early Skip/DIS</i> utilizando THR 90% . . . . .	76
Tabela 8	Resultados de simulação da heurística <i>Early Skip/DIS</i> utilizando THR 80% . . . . .	77
Tabela 9	Comparação entre os resultados obtidos com a heurística <i>early Skip/DIS</i> utilizando THR 99% e 95% . . . . .	78
Tabela 10	Comparação da heurística <i>Early Skip/DIS</i> com trabalhos relacionados	80

## LISTA DE ABREVIATURAS E SIGLAS

2D	Duas dimensões
3D	Três dimensões
ARP	<i>Advanced Residual Prediction</i>
AU	<i>Access Unit</i>
AVC	<i>Advanced Video Coding</i>
BD-rate	<i>Bjøntegaard Delta rate</i>
BSM	<i>Binary Segmentation Mask</i>
Cb	Amostra de Crominância Azul
CPV	<i>Constant Partition Value</i>
Cr	Amostra de Crominância Vermelha
csv	<i>Comma-Separated Value</i>
CSV	<i>Conventional Stereo Video</i>
CTC	<i>Common Test Conditions</i>
CTU	<i>Coding Tree Unit</i>
CU	<i>Coding Unit</i>
DBBP	<i>Depth-Based Block Partitioning</i>
DCP	<i>Disparity-compensated Prediction</i>
DF	<i>Deblocking Filter</i>
DIBR	<i>Depth Image Based Rendering</i>
DIS	<i>Depth Intra Skip</i>
dist	Distorção
DLT	<i>Depth Lookup Table</i>
DMM	<i>Depth Modelling Mode</i>
Do	<i>Depth-oriented</i>
DQP	<i>Depth Quadtree Prediction</i>
DvMCP	<i>Disparity Vector - Motion Compensated Prediction</i>

Gb	<i>Gigabit</i>
GB	<i>Gigabyte</i>
HD	<i>High Definition</i>
HEVC	<i>High Efficiency Video Coding</i>
HM	<i>HEVC Model</i>
HSI	<i>Hue, Saturation, Intensity</i>
HTM	<i>3D-HEVC Test Model</i>
IC	<i>Illumination Compensation</i>
IDE	<i>Integrated Development Environment</i>
IEC	<i>International Electrotechnical Commission</i>
IMP	<i>Interview Motion Prediction</i>
ISDC	<i>Intermode Segment-wise DC Coding</i>
ISO	<i>International Organization for Standardization</i>
IvpMv	<i>Interview Predicted Motion Vector</i>
JCT-VC	<i>Joint Collaborative Team on Video Coding</i>
MB	<i>Megabyte</i>
MC	<i>Motion Compensation</i>
MCL	<i>Merge Candidates List</i>
MCP	<i>Motion-compensated Prediction</i>
MD	<i>Mode Decision</i>
MPEG	<i>Moving Picture Expert Group</i>
MPI	<i>Motion Parameter Inheritance</i>
MV	<i>Moving Vector</i>
MVC	<i>Multiview Video Coding</i>
MVD	<i>Multiview plus Depth</i>
MVV	<i>Multiview Video</i>
NBDV	<i>Neighboring Block Disparity Vector</i>
PNG	<i>Portable Network Graphics</i>
Prof.	<i>Profundidade</i>
PU	<i>Prediction Unit</i>
QP	<i>Quantization Parameter</i>
QT	<i>Quadtree</i>
QTL	<i>Quadtree Limitation</i>
RAP	<i>Random Access Point</i>

RD	<i>Rate-Distortion</i>
RDC	<i>Rate-Distortion-Complexity</i>
Red.	Redução
RGB	<i>Red, Green, Blue</i>
RQT	<i>Residual Quadtree</i>
s	Segundo
SA	Sem Assimétricos
SAD	<i>Sum of Absolute Difference</i>
SAO	<i>Sample Adaptive Offset</i>
SDC	<i>Segment-wise DC Coding</i>
SR	Sem Retangulares
SSE	<i>Sum of Square Error</i>
Text.	Textura
TH	<i>Threshold</i>
THR	<i>Target Hit Rate</i>
TO	Taxa de Ocorrência
TOC	Taxa de Ocorrência Combinada
TU	<i>Transform Unit</i>
UHD	<i>Ultra-High Definition</i>
VCEG	<i>Video Coding Experts Group</i>
VDB	<i>Virtual Depth Block</i>
VSP	<i>View Synthesis Prediction</i>
Y	Amostra de Luminância
YCbCr	<i>Luminance, Chrominance Blue, Chrominance Red</i>

# SUMÁRIO

<b>1 INTRODUÇÃO . . . . .</b>	18
<b>1.1 Motivação . . . . .</b>	20
<b>1.2 Objetivo . . . . .</b>	21
<b>1.3 Organização do Trabalho . . . . .</b>	22
<b>2 VÍDEOS DIGITAIS E O PADRÃO HEVC . . . . .</b>	23
<b>2.1 Vídeos Digitais . . . . .</b>	23
2.1.1 Espaço de Cores e Subamostragem de Cores . . . . .	23
2.1.2 Redundância de Informações nos Vídeos . . . . .	25
<b>2.2 O Padrão HEVC . . . . .</b>	26
2.2.1 Estruturas de Bloco do HEVC . . . . .	27
2.2.2 Predição Intraquadro . . . . .	29
2.2.3 Predição Interquadros . . . . .	29
2.2.4 Modo <i>Merge</i> . . . . .	31
2.2.5 Modo <i>Skip</i> . . . . .	34
<b>2.3 Considerações Finais do Capítulo . . . . .</b>	35
<b>3 O PADRÃO 3D-HEVC . . . . .</b>	36
<b>3.1 Vídeos em três dimensões . . . . .</b>	36
<b>3.2 Sistema de Codificação do 3D-HEVC . . . . .</b>	37
<b>3.3 Ordem de Codificação do 3D-HEVC . . . . .</b>	38
<b>3.4 Codificação das Vistas Dependentes no 3D-HEVC . . . . .</b>	40
3.4.1 Predição Intervistas . . . . .	40
3.4.2 Obtenção do Vetor de Disparidade . . . . .	41
3.4.3 Refinamento do Vetor de Disparidade . . . . .	42
3.4.4 Modo <i>Merge</i> . . . . .	43
3.4.5 Predição de Movimento Intervistas . . . . .	43
3.4.6 Predição Avançada de Resíduos . . . . .	43
3.4.7 Compensação de Iluminação . . . . .	45
3.4.8 Predição Baseada em Síntese de Vistas . . . . .	45
3.4.9 Particionamento de Bloco Baseado em Profundidade . . . . .	45
<b>3.5 Codificação de Mapas de Profundidade no 3D-HEVC . . . . .</b>	45
3.5.1 Modos Intra para Modelagem de Profundidade . . . . .	46
3.5.2 Codificação <i>Segment-wise DC</i> . . . . .	48
3.5.3 Herança de Parâmetros de Movimento . . . . .	48
3.5.4 Codificação <i>Segment-wise DC</i> no Modo <i>Inter</i> . . . . .	49
3.5.5 Predição de <i>Quadtree</i> de Profundidade . . . . .	49

3.5.6 Modo <i>Skip Intra</i> de Profundidade . . . . .	49
<b>3.6 Considerações Finais do Capítulo . . . . .</b>	<b>51</b>
<b>4 ANÁLISES SOBRE A CODIFICAÇÃO DOS MAPAS PROFUNDIDADE NO 3D-HEVC . . . . .</b>	<b>52</b>
4.1 Análise de distribuição dos modos de predição . . . . .	53
4.2 Análise RDC com a remoção de modos de predição . . . . .	57
4.3 Considerações Finais do Capítulo . . . . .	62
<b>5 HEURÍSTICA <i>EARLY SKIP/DIS</i> . . . . .</b>	<b>64</b>
5.1 Análise de ocorrência dos modos <i>Skip</i> e <i>DIS</i> em função do custo RD . . . . .	65
5.2 Detalhamento da heurística <i>Early Skip/DIS</i> . . . . .	68
5.2.1 Obtenção dos <i>Thresholds</i> . . . . .	69
5.2.2 Algoritmo do núcleo da heurística <i>early Skip/DIS</i> . . . . .	72
5.3 Resultados da heurística <i>Early Skip/DIS</i> . . . . .	72
5.4 Trabalhos Relacionados e Comparações . . . . .	78
5.5 Considerações Finais do Capítulo . . . . .	80
<b>6 CONCLUSÃO . . . . .</b>	<b>82</b>
<b>REFERÊNCIAS . . . . .</b>	<b>85</b>
<b>APÊNDICE A FERRAMENTA 3D-HEVC ANALYZER . . . . .</b>	<b>88</b>
<b>APÊNDICE B TRACER HEVC/3D-HEVC . . . . .</b>	<b>93</b>

# 1 INTRODUÇÃO

Atualmente está ocorrendo uma popularização gradual de dispositivos capazes de capturar e reproduzir vídeos digitais em três dimensões (3D), tais como televisores, *Blu-ray players*, *tablets*, *smartphones*, etc. Este tipo de mídia permite aos espectadores uma experiência visual mais qualificada a partir da percepção de distância entre a câmera e os objetos da cena. Isto é possível a partir do uso de *displays* especiais, os quais possibilitam que a cena seja projetada ao olho esquerdo e direito do espectador com uma diferença de deslocamento horizontal similar ao percebido no sistema visual humano (TECH et al., 2015). Apesar do acréscimo na experiência visual proporcionado, o armazenamento e o processamento de vídeos 3D possuem diversos desafios associados.

O primeiro desafio destacado é relacionado ao volume de bits necessário para representar um vídeo 3D sem codificação. Como exemplo, considerando um vídeo com duas vistas (mínimo em um sistema 3D), com cada vista em resolução HD 1080p (1920x1080 *pixels*) e que utilize três bytes para representar numericamente cada pixel (três amostras de cores por *pixel* com um byte por amostra), um quadro deste vídeo em cada uma das vistas é composto por 2.073.600 *pixels*, utilizando 6.220.800 bytes em sua representação. Desta forma, considerando uma taxa de exibição de 30 quadros por segundo, a largura de banda necessária seria de 373,3 MB/s (2,99 Gb/s) caso este vídeo fosse transmitido pela internet sem nenhuma técnica de compressão. Considerando este vídeo com apenas 10 minutos de duração, seria necessário um espaço em disco de aproximadamente 223 GB para o seu armazenamento. Estes números de largura de banda para transmissão e capacidade de armazenamento demonstram a inviabilidade da manipulação de vídeos 3D sem compressão.

O *High Efficiency Video Coding* (HEVC) (ITU-T, 2014) é o mais novo padrão de codificação de vídeos e foi desenvolvido pela junção de *experts* da *International Organization for Standardization/International Electrotechnical Commission* (ISO/IEC) *Moving Picture Expert Group* (MPEG) e da *International Telecommunication Union* (ITU) *Telecommunication Standardization Sector* (ITU-T) *Video Coding Experts Group* (VCEG), formando então o *Joint Collaborative Team on Video Coding* (JCT-VC, 2010),

sendo lançado em janeiro de 2013. Entre as principais metas do HEVC quando fora proposto, destacava-se atingir o dobro da taxa de compressão - quando comparado ao seu antecessor, o padrão H.264/AVC (RICHARDSON, 2003) - mantendo a mesma qualidade de imagem e complexidade computacional (SULLIVAN et al., 2012). Além disso, esperava-se que o HEVC apresentasse suporte à codificação eficiente de vídeos de resolução até UHD 8K (7680x4320 pixels), 16 vezes maior que a resolução HD 1080p para qual seu predecessor, H.264/AVC, fora desenvolvido (WIEGAND et al., 2003). Devido ao número de ferramentas e estruturas de codificação utilizadas para obter os excelentes resultados de compressão de vídeo, não foi possível ao HEVC atingir sua meta de manter a mesma complexidade computacional do H.264/AVC. Assim, o esforço computacional para o HEVC codificar um vídeo chega a ser 502% maior que o H.264/AVC (CORREA et al., 2012).

A fim de expandir a eficiência de codificação a diversos outros tipos de aplicações, extensões para os padrões atuais de codificação de vídeos foram desenvolvidos visando dar suporte a vídeos capturados por mais de uma câmera simultaneamente. O *Multiview Video Coding* (MVC) (JVT, 2009) foi o padrão de codificação de vídeos multivista desenvolvido com base no padrão H.264/AVC. Com base no HEVC, experts desenvolveram o *3D-High Efficiency Video Coding* (3D-HEVC) (TECH et al., 2015). O 3D-HEVC foi desenvolvido visando codificar eficientemente vídeos 3D, levando em conta as características específicas deste tipo de aplicação. A complexidade computacional apontada no HEVC cresce significativamente quando considerado o processo de codificação de vídeos 3D, visto que o número de imagens processadas para um mesmo instante de tempo (vistas) é maior. Além das redundâncias exploradas pelo codificador convencional, o 3D-HEVC também explora redundâncias entre imagens pertencentes ao mesmo instante de tempo, elevando ainda mais a complexidade computacional do 3D-HEVC.

Somando-se a isto, o 3D-HEVC introduziu o conceito de mapas de profundidade (*depth maps*) (FEHN et al., 2006) na codificação de vídeos, que são imagens monocromáticas onde informações de distância dos objetos da cena em relação à câmera são armazenadas. Mapas de profundidade são essenciais no processo de síntese de vistas, o qual, a partir de dois quadros de textura (imagem convencional) capturados em posições diferentes (duas vistas), juntamente com seus respectivos mapas de profundidade, possibilitam a geração de uma vista intermediária entre as duas câmeras - denominada vista sintética (MULLER et al., 2013). Desta forma, além da complexidade adicional em relação ao HEVC, devido a exploração das redundâncias entre as imagens capturadas durante o mesmo instante de tempo, o codificador 3D-HEVC também apresenta um acréscimo na complexidade ao processar mapas de profundidade.

Devido ao elevado número de possibilidades de particionamento de blocos em uni-

dades de codificação, unidades de predição e unidades de transformada; somado aos diversos modos e ferramentas de codificação que podem ser empregados no codificador HEVC; o processo de codificação de vídeos torna-se extremamente complexo, exigindo um esforço computacional elevado. Estudos realizados em trabalhos relacionados utilizando o software de referência do HEVC, o *HEVC Model* (HM) versão 8.0 (MC-CANN et al., 2012) demonstram que o tempo de codificação do HEVC, para um vídeo 2D com resolução HD 1080p, sendo efetuada em um *cluster* Xeon E5670 (2,93 GHz), pode exceder em até 4.000 vezes o tempo necessário exigido em aplicações de tempo real (BOSSEN et al., 2012). Além disso, considerando o acréscimo significativo no volume de dados a serem processados nos vídeos 3D, além do aumento no número de ferramentas utilizadas pelo 3D-HEVC, este problema torna-se ainda mais crítico no processo de codificação de vídeos em três dimensões.

## 1.1 Motivação

Desta forma, justifica-se a necessidade de estratégias inteligentes que busquem reduzir a complexidade computacional do processo de codificação de vídeos do 3D-HEVC. Estas estratégias não devem gerar um *bitstream* incompatível com o padrão, possibilitando a integração da solução em dispositivos que capturem e codifiquem o vídeo de forma compatível com o 3D-HEVC. Além disso, estas soluções de redução de complexidade não podem apresentar impacto significativo na eficiência da codificação, visando não comprometer a qualidade e a taxa de compressão obtidas pelo codificador.

Estratégias propostas neste cenário geralmente são inseridas no módulo de decisão de modo (*Mode Decision* - MD) do codificador, onde as diversas possibilidades de particionamento de blocos em unidades de codificação, predição e transformada são avaliados; além dos diversos modos e ferramentas disponíveis. A Figura 1 apresenta um diagrama de blocos de alto nível para o 3D-HEVC, mais especificamente para o codificador de mapas de profundidade deste padrão. Cada imagem processada pelo codificador de mapas de profundidade do 3D-HEVC é avaliada pelos modos de predição intraquadro (blocos cinzas) e interquadros (blocos brancos). O resultado da predição é subtraído da informação original, gerando os blocos residuais. Cada bloco residual é transformado e quantizado, sendo então enviado ao codificador de entropia. Visto que a quantização insere erros irreversíveis nas informações residuais, o bloco atual é reconstruído para ser utilizado como referência para a predição dos demais blocos, evitando discrepâncias entre as referências utilizadas pelo codificador e decodificador (SALOMON, 2006).

Como é possível perceber, o MD está diretamente relacionado a diversos módulos do codificador, tais como o *Codificador de Entropia*; a codificação residual, composta

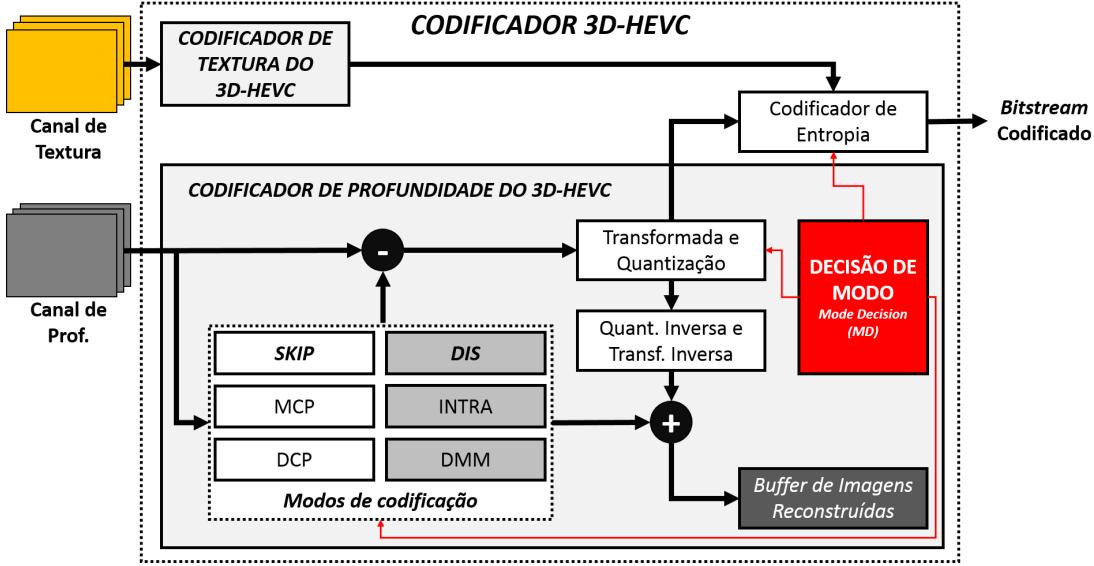


Figura 1: Diagrama de blocos de alto nível para o codificador de vídeos 3D-HEVC

pelos módulos de *Transformada* e *Quantização*; e por fim, também está ligado diretamente à aplicação dos modos de codificação. Na Figura 1, alguns modos de codificação que exploram a correlação temporal (blocos brancos) foram listados: o modo *Skip*, a Predição Compensada pelo Movimento (*Motion-Compensated Prediction* - MCP) e a Predição Compensada pela Disparidade (*Disparity-Compensated Prediction* - DCP). Somando-se a estes, alguns modos que exploram a correlação espacial também foram listados: o *Skip Intra* de Profundidade (*Depth Intra Skip* - DIS), a Predição Intraquadro convencional (INTRA) e os Modos Intra para Modelagem de Profundidade (*Depth Modelling Modes* - DMM). Todos estes modos, além de outras ferramentas de codificação, serão abordados nas seções seguintes deste trabalho.

## 1.2 Objetivo

Esta dissertação tem por objetivo propor uma solução heurística para o módulo MD do 3D-HEVC, a qual visa reduzir a complexidade computacional do codificador de mapas de profundidade, sem apresentar impactos significativos na eficiência de codificação do padrão. A heurística proposta visa testar, primeiramente, os modos *Skip* e DIS do codificador de mapas de profundidade do 3D-HEVC e, dado o custo de codificação que relaciona a taxa de bits e qualidade de imagem (*Rate-Distortion* - RD) destes modos, é tomada a decisão de avaliar ou não as demais ferramentas de codificação. Dado que esta solução visa selecionar antecipadamente a utilização dos modos *Skip* e/ou DIS, ela é nomeada *Early Skip/DIS*. Os valores limiares (*thresholds*) que determinarão se os demais modos serão ou não avaliados são definidos durante a codificação dos primeiros quadros do vídeo, ou em intervalos pré-definidos, visando obter uma pré-determinada taxa de acerto. Assim, a definição destes valores é reali-

zada de forma dinâmica e adaptativa para cada vídeo e configurações de codificação utilizadas, aprimorando a eficiência de redução de complexidade da heurística *Early Skip/DIS*.

O desenvolvimento desta heurística é conduzido através de uma análise estatística inicial, que demonstra quais são os principais modos de particionamento e ferramentas de codificação de mapas de profundidade empregados no 3D-HEVC. Além disso, esta análise também apresenta a relação entre taxa de bits, qualidade de imagem e complexidade (*Rate-Distortion-Complexity* - RDC) ao desabilitar determinados dos modos de particionamento das unidades de predição que apresentam baixa ocorrência.

### 1.3 Organização do Trabalho

Este trabalho está dividido da seguinte maneira. O Capítulo 2 apresentará uma revisão bibliográfica sobre vídeos digitais e o codificador HEVC, onde modos e ferramentas de codificação deste padrão serão apresentadas. O Capítulo 3 é destinado a descrever o processo de codificação de vídeos em três dimensões, realizado pelo 3D-HEVC. As ferramentas e modos de codificação, que exploram as redundâncias não abordadas pelo HEVC, são brevemente descritas neste capítulo. A seguir, o Capítulo 4 descreve a análise RDC que avalia a relevância dos modos de particionamento do codificador de mapas de profundidade do 3D-HEVC. O Capítulo 5 é destinado a abordar a heurística *Early Skip/DIS* desenvolvida, descrevendo os algoritmos de geração dos *thresholds*, tal como as mudanças efetuadas no módulo da MD. Resultados de simulações das estratégias desenvolvidas são apresentadas no Capítulo 5. Somando-se a isto, comparações com trabalhos relacionados também serão comentadas. Por fim, o Capítulo 6 descreve as considerações finais deste trabalho, apontando os próximos passos a serem seguidos em trabalhos futuros.

## 2 VÍDEOS DIGITAIS E O PADRÃO HEVC

Este capítulo primeiramente apresenta os conceitos básicos de vídeos digitais, juntamente com os conceitos de espaço de cores, subamostragem e redundância de dados presentes nos vídeos digitais. Posteriormente, este capítulo apresentará, de forma resumida, as ferramentas e estruturas de codificação empregadas pelo HEVC, as quais foram herdadas pelo 3D-HEVC, discutido no próximo capítulo.

### 2.1 Vídeos Digitais

Um vídeo é basicamente uma sequência de imagens, ou também quadros (frames), que aparecem em uma determinada frequência a fim de dar ao espectador a sensação de movimento contínuo. Para isto, é necessário que a frequência mínima de exibição destas imagens seja de 24 a 30 quadros por segundo (GONZALEZ; WOODS, 2008).

Ao armazenar e/ou transmitir a informação na forma digital, as informações são basicamente convertidas para uma sequência binária (zeros e uns). Esta abordagem, embora apresente uma determinada perda na qualidade devido à amostragem realizada ao digitalizar as informações, apresenta diversos benefícios quando comparado ao formato analógico. Entre estes benefícios, destacam-se o melhor aproveitamento do canal de transmissão e a baixa sensibilidade a erros, provenientes de ruídos presentes no canal de transmissão (LATHI, 1998).

#### 2.1.1 Espaço de Cores e Subamostragem de Cores

Encontram-se na literatura diversas formas de representar imagens coloridas em formato digital, como por exemplo: RGB, HSI e YCbCr (SHI; SUN, 2008). O sistema pelo qual as cores são representadas é denominado espaço de cores. Dentre os citados, o RGB é um dos mais comuns e conhecidos. O RGB representa, em três matrizes distintas, as três cores primárias (considerando combinações aditivas de cores): vermelho, verde e azul. Com base nisto, surge o nome deste espaço de cores (do inglês *red green blue* - RGB) (AGOSTINI, 2007).

No espaço de cores YCbCr, as três componentes utilizadas são: luminância (Y), crominância azul (Cb), e crominância vermelha (Cr). Todas as cores representadas no espaço RGB são possíveis ser representadas no espaço YCbCr. Os componentes R, G e B do espaço de cores RGB possuem um elevado grau de correlação e isto torna difícil o processamento de cada uma das informações de cor de forma independente (RICHARDSON, 2002). Este é o principal fator que faz com que os padrões de compressão de vídeo utilizem espaço de cores do tipo luminância e crominância, como o YCbCr.

Outro fator que justifica o uso do espaço de cores YCbCr é que o sistema visual humano é mais sensível a informações de luminância do que a informações de crominância. Desta forma, os padrões de compressão de vídeos podem explorar esta característica humana para aumentar a eficiência da codificação através da redução da taxa de amostragem dos componentes de crominância em relação aos componentes de luminância (RICHARDSON, 2002). A redução das informações de crominância é chamada de subamostragem de cores, sendo realizada no espaço de cores YCbCr nos padrões de compressão de vídeos atuais.(AGOSTINI, 2007).

Existem diversos formatos de subamostragem para os componentes de crominância. Dentre estes, destacam-se os formatos 4:4:4, 4:2:2 e 4:2:0. No formato 4:4:4, para cada amostra de luminância (Y), existe uma amostra de crominância azul (Cb) e uma amostra de crominância vermelha (Cr) associada; ou seja, as amostras Cb e Cr não são subamostradas. No formato 4:2:2, a cada duas amostras de luminância (Y), são associadas uma amostra de crominância azul (Cb) e uma amostra de crominância vermelha (Cr). Por último, no formato 4:2:0, para cada quatro amostras de luminância (Y), existem apenas uma amostra de crominância azul (Cb) e uma amostra de crominância vermelha (Cr). A nomenclatura deste último formato não representa uma relação lógica entre as componentes de crominância, sendo usada por motivos históricos (RICHARDSON, 2003).

A subamostragem de cores aumenta significativamente a eficiência do processo de codificação de vídeos, visto que este método elimina boa parte das informações de crominância (AGOSTINI, 2007). Considerando como exemplo um vídeo que utilize subamostragem no formato 4:2:0, é necessário apenas um quarto de informação para cada canal de crominância comparado com o canal de luminância. Assim, para cada quatro amostras de luminância, são utilizadas mais uma amostra de crominância azul e uma de crominância vermelha, resultando em seis amostras apenas. Desta forma, atinge-se uma taxa de compressão de 50% sem perdas perceptíveis ao sistema visual humano, quando comparado ao formato 4:4:4 (AGOSTINI, 2007).

### 2.1.2 Redundância de Informações nos Vídeos

A codificação de vídeos busca diminuir a quantidade de dados que não acrescentam informações novas ao vídeo; ou seja, dados redundantes na representação do mesmo (AGOSTINI, 2007). Na codificação de vídeos, basicamente três diferentes tipos de redundância são explorados no processo de compressão: a redundância espacial, a redundância temporal e a redundância entrópica (AGOSTINI, 2007).

A redundância espacial, também conhecida como redundância intraquadro (GHANBARI, 2003), refere-se à correlação existente entre *pixels* espacialmente distribuídos em um quadro de um dado vídeo ou imagem estática. Esta correlação pode ser percebida, tanto no domínio espacial, quanto no domínio das frequências (AGOSTINI, 2007). No domínio espacial, a correlação entre os *pixels* pode ser percebida de forma visual, onde normalmente os *pixels* vizinhos tendem a ser semelhantes. Neste caso, a redundância pode ser reduzida através da operação chamada predição intraquadro, presente em vários padrões de codificação de vídeos atuais (AGOSTINI, 2007). No domínio das frequências, a redundância espacial é reduzida através da quantização. Antes de serem quantizadas, as informações devem ser transformadas do domínio espacial para o domínio das frequências. A quantização é basicamente uma divisão inteira dos coeficientes gerados pela transformada e reduz grande parte dos coeficientes a zero. Esta operação é irreversível, pois o resto da divisão não é armazenado (AGOSTINI, 2007). Desta forma, o processo de remover redundâncias espaciais no domínio das frequências gera perdas de qualidade no vídeo codificado. Entretanto, estas perdas tendem a interferir de forma pouco significativa na qualidade perceptual da imagem, visto que as atenuações são realizadas de forma mais agressiva em frequências menos perceptíveis ao sistema visual humano (AGOSTINI, 2007).

A redundância temporal, também conhecida como redundância interquadros (GHANBARI, 2003), é dada pela correlação existente entre quadros temporalmente vizinhos em um vídeo (AGOSTINI, 2007). Muitos blocos de *pixels* não mudam de valor entre um quadro e outro; ou, em outros casos, os blocos apenas se movimentam em relação aos quadros vizinhos. Por exemplo, o movimento de um objeto em uma dada cena (AGOSTINI, 2007). Os padrões atuais de codificação de vídeos visam reduzir a redundância temporal de forma eficiente, conduzindo à elevadas taxas de compressão, o que é fundamental para o sucesso dos codificadores de vídeo (AGOSTINI, 2007).

Por último, existe a redundância entrópica, a qual está relacionada com as probabilidades de ocorrência dos símbolos codificados (AGOSTINI, 2007). A entropia é uma medida da quantidade média de informação transmitida por símbolo do vídeo (SHI; SUN, 2008). Quanto maior a probabilidade de ocorrência de um símbolo, menor a quantidade de informação nova transmitida (AGOSTINI, 2007).

Todas estas redundâncias são amplamente exploradas pelos padrões atuais de codificação de vídeo, visando reduzir drasticamente a quantidade de bits necessários na representação dos mesmos.

## 2.2 O Padrão HEVC

Esta seção apresenta uma revisão bibliográfica resumida das ferramentas e estruturas empregadas pelo HEVC. As referências base utilizadas foram: o *Overview* do HEVC (SULLIVAN et al., 2012), o livro que apresenta algoritmos e arquiteturas de hardware para o HEVC (SZE; BUDAGAVI; SULLIVAN, 2014) e o *Test Model 16* do HEVC (ROSEWARNE et al., 2015).

O HEVC apresenta uma abordagem de codificação híbrida (predições interquadros e intraquadro seguidos por uma transformada 2D e um codificador de entropia), assim como diversos outros padrões de codificação. Entretanto, diferentemente de padrões anteriores, o HEVC emprega uma estrutura de particionamento de blocos em árvore quaternária (*quadtree - QT*), possibilitando a seleção flexível dos tamanhos de blocos de predição, transformada e codificação (SULLIVAN et al., 2012). Esta flexibilidade nas estruturas é um dos principais fatores que contribuem para a obtenção dos eficientes resultados de compressão obtidos pelo HEVC, quando comparado aos demais padrões.

A Figura 2 apresenta um diagrama de blocos de um codificador de vídeos, o qual é capaz de gerar um *bitstream* de acordo com as normas de padronização do HEVC. Cada quadro do vídeo é dividido em blocos denominados *Coding Tree Units* (CTUs), os quais serão explicados de forma mais ampla a seguir. O primeiro quadro de um vídeo é codificado apenas com predição intraquadro, a qual utiliza dados da própria imagem para realizar a predição. Esta predição não gera dependências do quadro atual com os demais quadros (SULLIVAN et al., 2012).

Para os demais quadros de uma sequência de vídeo, a codificação interquadros é utilizada para a maioria dos blocos. O processo da predição interquadros consiste em escolher dados de movimento, os quais compreendem o quadro de referência e o vetor de movimento (*Moving Vector - MV*). O codificador e o decodificador geram os mesmos sinais de predição interquadros através da aplicação da compensação de movimento (*Motion Compensation - MC*), utilizando o MV e os dados da decisão de modo, os quais são transmitidos como informações laterais pelo codificador.

O resíduo da predição intraquadro e interquadros, que é a diferença entre o bloco original e o predito, é transformado do domínio espacial para o domínio das frequências, utilizando uma transformada espacial linear. O bloco transformado é, então, escalado, quantizado, codificado entropicamente e transmitido juntamente com as informações de predição. Visto que a codificação residual insere erros no processo

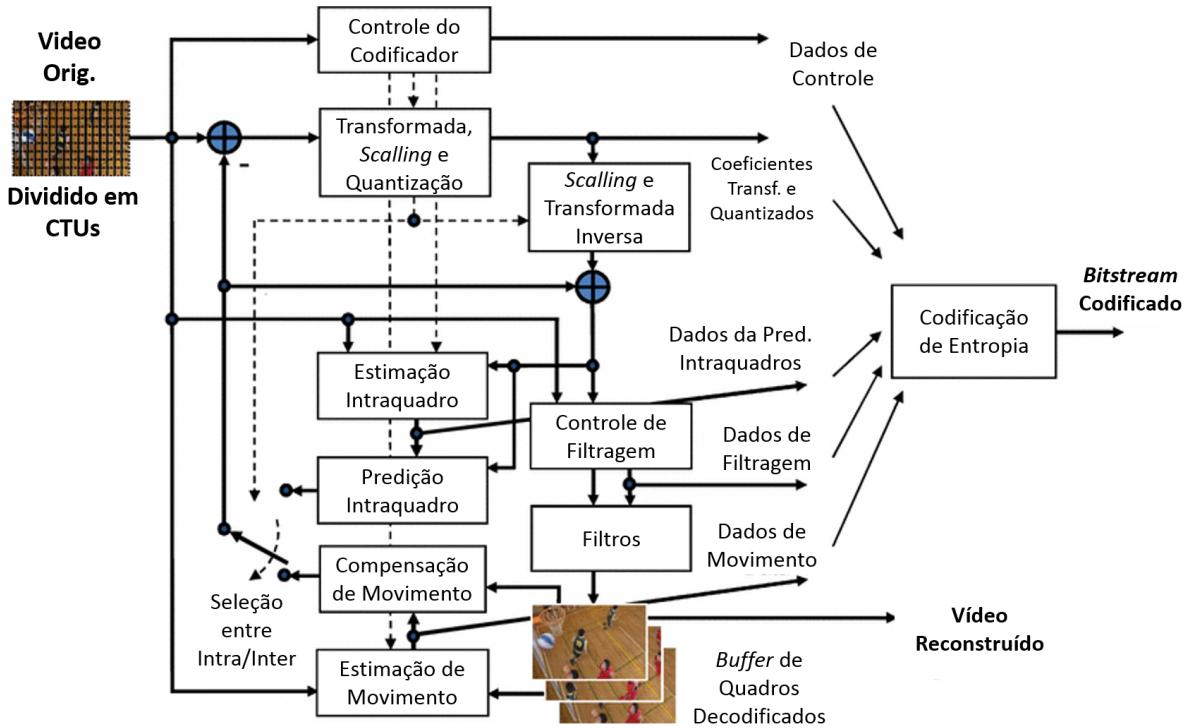


Figura 2: Diagrama de blocos do codificador de vídeos HEVC. Adaptada de (SULLIVAN et al., 2012)

de codificação, o codificador também implementa o decodificador (caixas cinzas na Figura 2). Desta forma, é possível que ambos gerem previsões idênticas para as informações subsequentes, evitando erros entre os dados processados pelo codificador e pelo decodificador.

### 2.2.1 Estruturas de Bloco do HEVC

O núcleo da camada de codificação no HEVC é a CTU (SULLIVAN et al., 2012), a qual é composta por um bloco de luminância e dois blocos correspondentes de crominância. O tamanho de uma CTU é definido pelo tamanho do bloco de luminância, que pode ser 16x16, 32x32 ou 64x64 (ROSEWARNE et al., 2015). Visto que o número de aplicações que utilizam CTUs de tamanho 16x16 e 32x32 é muito restrito, neste trabalho será considerado que CTUs são exclusivamente de tamanho 64x64.

Os atributos da CTU especificam o tamanho e a localização das *Coding Units* (CUs), as quais possuem um bloco de luminância - cujo tamanho define o tamanho da CU - e dois blocos correspondentes de crominância, além dos atributos específicos da unidade (SZE; BUDAGAVI; SULLIVAN, 2014). Uma vez que a raiz de uma *quadtree* é associada a uma CTU, o tamanho de uma CTU é o maior tamanho possível de uma CU. Cada CTU pode conter apenas uma CU ou ser dividida em quatro CUs de mesmo tamanho. Além disso, cada CU pode ser dividida recursivamente até que a CU folha seja de tamanho 8x8. Por fim, cada CU contém um particionamento em *Prediction*

*Units* (PUs) e uma árvore de *Transform Units* (TUs).

Cada CU folha na *quadtree* é associada a uma ou mais PUs de acordo com o modo de partição, e todas PUs associadas a uma dada CU utilizam o mesmo modo de predição (intraquadro ou interquadros) (ROSEWARNE et al., 2015). Cada CU possui uma, duas ou quatro PUs, de acordo com os oito possíveis modos de particionamento de CU em PUs, conforme apresentado na Figura 3.

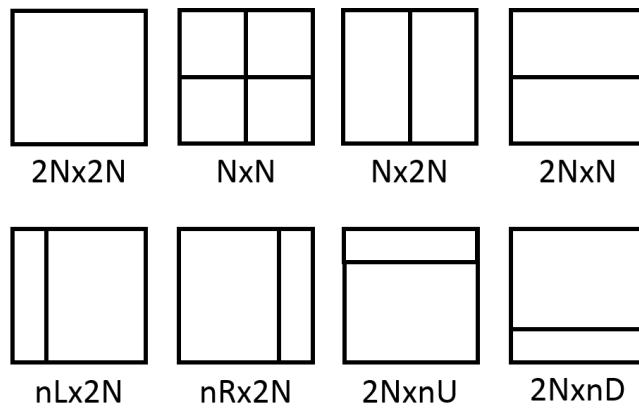


Figura 3: Particionamento de CUs em PUs

Quando uma CU é definida com predição intraquadro, apenas os particionamentos quadrados são habilitados ( $2Nx2N$  e  $NxN$ ), sendo que o menor destes somente está disponível no caso em que o tamanho da CU é o menor possível ( $8x8$ ) (ROSEWARNE et al., 2015). Assim, cada PU neste caso poderá ter o tamanho de  $4x4$  amostras. Para uma CU definida com predição interquadros, todos os oito modos de particionamento estão disponíveis.

Uma das inovações do HEVC frente aos demais padrões de codificação de vídeos foi o uso de estruturas de predição retangulares (assimétricas). Esta técnica permite que o codificador efetue uma estimativa de movimento mais eficiente em bordas reais de objetos da imagem. Embora esta técnica eleve a eficiência de codificação do HEVC, o uso de PUs assimétricas eleva a complexidade computacional do processo devido ao acréscimo no número de modos a serem testados.

Para a codificação residual, uma CU pode ser recursivamente particionada em TUs, formando a *Residual Quadtree* (RQT). Cada TU pode ser de tamanho  $32x32$ ,  $16x16$ ,  $8x8$  e  $4x4$ . Caso a CU folha seja de tamanho  $64x64$ , esta então conterá quatro RQTs de tamanho  $32x32$  na raiz (SZE; BUDAGAVI; SULLIVAN, 2014). A Figura 4 apresenta um exemplo de subdivisão de uma CTU em CUs e subdivisão das CUs em TUs. As linhas contínuas representam bordas de CUs ao passo que linhas pontilhadas representam bordas de TUs.

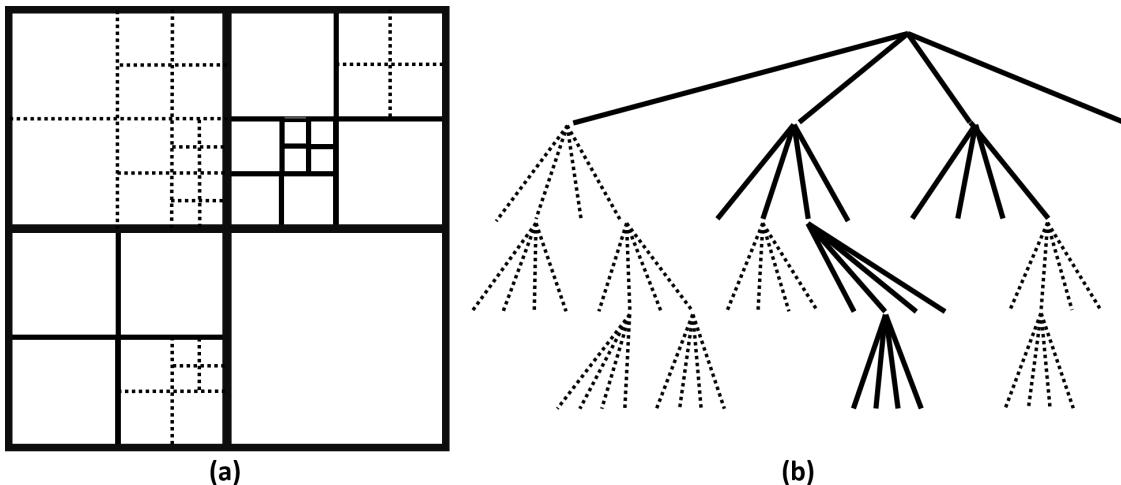


Figura 4: Subdivisão de uma CTU em CUs (linhas contínuas) juntamente subdivisão de CUs folhas em TUs (linhas pontinhadas). (a) CTU e suas subdivisões internas. (b) Quadtree + RQT correspondente

### **2.2.2 Predição Intraquadro**

A predição intraquadro é uma ferramenta de codificação que visa remover redundâncias espaciais utilizando amostras vizinhas previamente codificadas dentro de uma imagem. O HEVC estipula 35 diferentes modos de predição intraquadro, sendo 33 destes direcionais (angulares), um planar e um DC (SULLIVAN et al., 2012). A Figura 5 apresenta os ângulos dos modos direcionais de predição intraquadro do HEVC ao passo que a Figura 6 mapeia cada direção da Figura 5 para os modos de predição intraquadro do HEVC.

A Figura 7 exemplifica o funcionamento do modo 26 da predição intraquadro, onde os blocos em cinza representam amostras previamente codificadas da PU acima e os blocos brancos representam as amostras a serem preditas. Conforme pode ser observado na Figura 6, o modo 26 é um modo vertical, ou seja, cada amostra a ser predita é proveniente da amostra acima, o que não acontece em modos de predição diagonais, onde são aplicados filtros a fim de serem geradas as amostras.

### 2.2.3 Predição Interquadros

Enquanto a predição intraquadro explora a correlação espacial entre amostras vizinhas, a predição interquadros faz uso da correlação temporal entre imagens, visando obter uma predição compensada pelo movimento (*Motion-compensated Prediction - MCP*) para cada PU interquadros.

O conceito geral da MCP é baseado no modelo de movimento translacional (SZE; BUDAGAVI; SULLIVAN, 2014), conforme ilustrado na Figura 8. Utilizando este modelo, a posição do bloco atual nos quadros previamente codificados é indicada por um vetor de movimento ( $\Delta_x, \Delta_y$ ), onde  $\Delta_x$  indica o deslocamento horizontal e  $\Delta_y$  indica

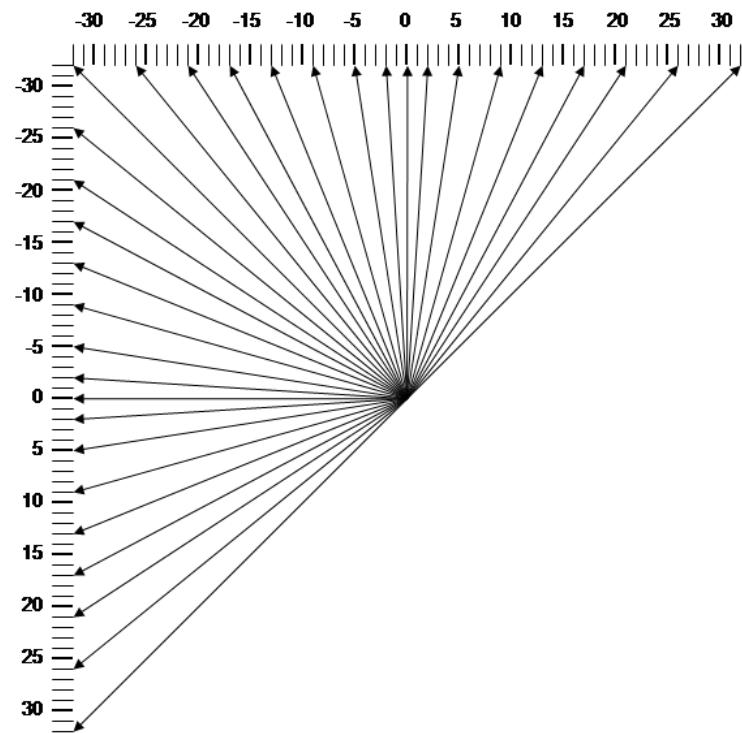


Figura 5: 33 modos direcionais e seus respectivos ângulos da predição intraquadro (ROSEWARNE et al., 2015)

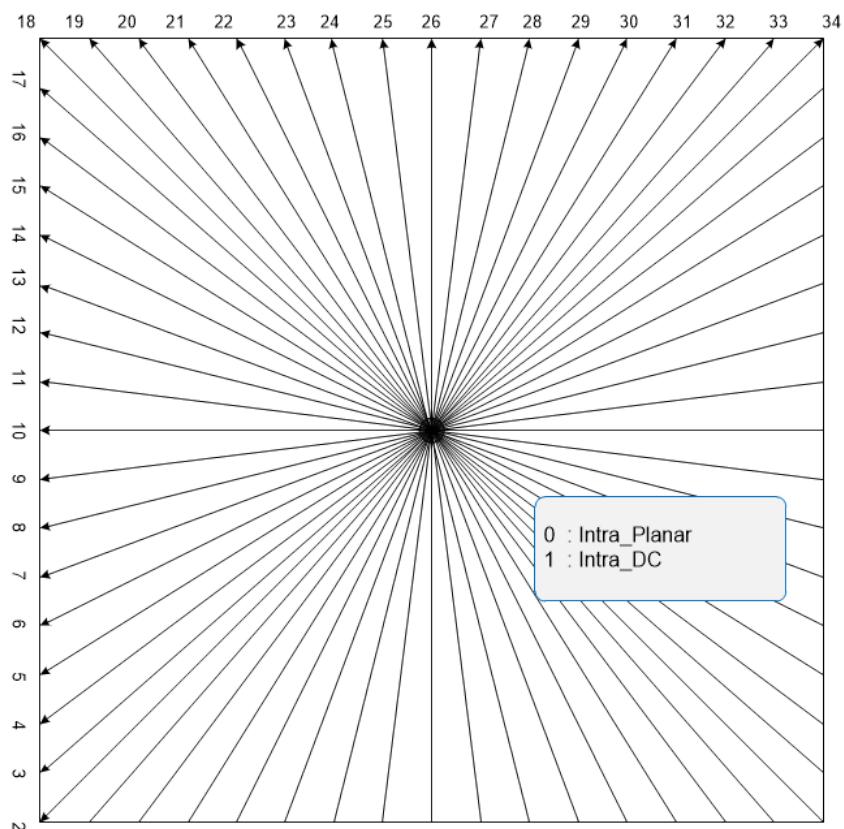


Figura 6: Mapeamento entre as direções da predição intraquadro e os modos de predição (ROSEWARNE et al., 2015)

A	B	C	D
$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$
$X_{21}$	$X_{22}$	$X_{23}$	$X_{24}$
$X_{31}$	$X_{32}$	$X_{33}$	$X_{34}$
$X_{41}$	$X_{42}$	$X_{43}$	$X_{44}$

**(a)**

A	B	C	D
A	B	C	D
A	B	C	D
A	B	C	D

**(b)**

Figura 7: Exemplo de aplicação do modo 26 de predição intraquadro. (a) bloco original (b) bloco predito

o deslocamento vertical do bloco previamente codificado em relação ao bloco atual. Vetores de movimento não são restritos a valores inteiros, podendo assumir valores fracionários a fim de capturar mais precisamente a posição dos objetos entre diferentes quadros subjacentes. Para isto, um filtro de interpolação é aplicado no quadro previamente codificado, visando derivar as amostras virtuais intermediárias, possibilitando que o vetor de movimento possua precisão fracionária (SULLIVAN et al., 2012).

O quadro previamente codificado é chamado de quadro de referência e é indicado por um índice de referência  $\Delta t$  em uma lista de quadros de referência, que, juntamente com o vetor de movimento, formam os dados de movimento.

Duas formas de predição interquadros estão disponíveis no HEVC, assim como em diversos padrões de codificação de vídeo atuais, denominados uni-predição e bi-predição. Uni-predição ocorre quando é utilizado apenas um conjunto de dados de movimento ( $\Delta x, \Delta y, \Delta t$ ) na geração da MCP. No caso da bi-predição, dois conjuntos de dados de movimento ( $\Delta x_0, \Delta y_0, \Delta t_0$  e  $\Delta x_1, \Delta y_1, \Delta t_1$ ) são utilizados para gerar duas MCPs, que podem referenciar dois quadros diferentes. Quando combinadas, estas duas MCPs geram uma MCP final (SZE; BUDAGAVI; SULLIVAN, 2014).

Além disso, outros modos, como o modo *Merge* e o modo *Skip* também são avaliados, conforme descrito a seguir.

#### 2.2.4 Modo *Merge*

Em codificação de vídeos, o particionamento da imagem em blocos é uma abordagem eficiente e comumente utilizada, visto que possibilita a aplicação de diferentes modos de predição, com níveis de esforço distintos, para regiões com maior/menor nível de detalhamento. Entretanto, devido às restrições de tamanho de bloco, este particionamento pode separar duas regiões da imagem que possuem mesmas características e que possivelmente seriam codificados com mesmo modo de

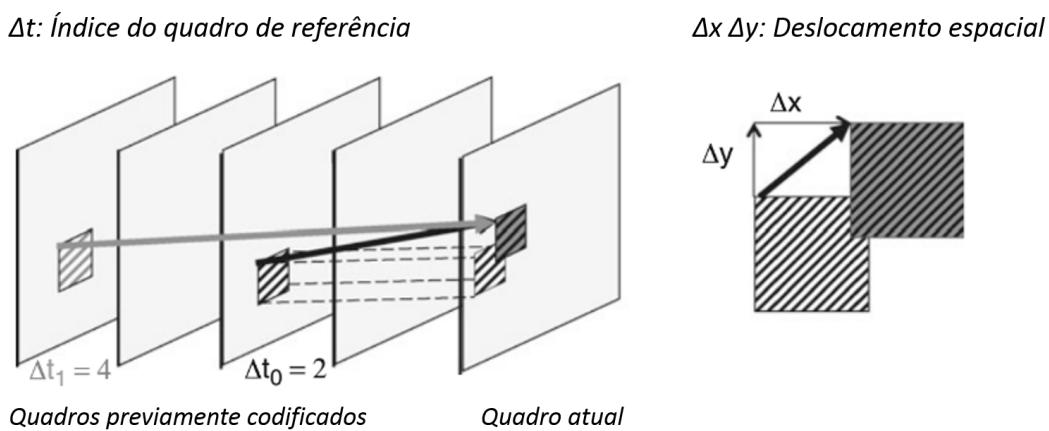


Figura 8: Funcionamento da predição interquadros do HEVC. Adaptada de (SZE; BUDAGAVI; SULLIVAN, 2014)

predição (SZE; BUDAGAVI; SULLIVAN, 2014).

Visando minimizar estes inconvenientes gerados pela codificação baseada em blocos, o HEVC definiu o modo *Merge*, o qual possibilita que dados de movimento sejam herdados de blocos vizinhos (espacial e temporal).

O primeiro passo realizado no modo *Merge* do HEVC é a definição dos blocos candidatos para derivar os dados de movimento, visando gerar a lista de candidatos *Merge* (*Merge Candidates List* - MCL). Até quatro blocos candidatos para o modo *Merge* são selecionados, dentre cinco possíveis blocos vizinhos espaciais, e um bloco candidato é selecionado dentre dois possíveis blocos co-localizados temporalmente. É importante salientar que este processo de geração da MCL é realizado tanto no codificador, quanto no decodificador, visando reduzir a quantidade de informação transmitida (ROSEWARNE et al., 2015).

A Figura 9 apresenta os vizinhos espaciais avaliados durante a construção da MCL. A ordem da avaliação é  $A_1$ ,  $B_1$ ,  $A_0$ ,  $B_0$  e  $B_2$ . A posição  $B_2$  somente é considerada quando qualquer PU na posição  $A_1$ ,  $B_1$ ,  $A_0$  ou  $B_0$  não estiver disponível ou for codificada com predição intraquadro. Após a PU na posição  $A_1$  ser adicionada na MCL, uma verificação de redundância é realizada, garantindo que candidatos com mesmos dados de movimento serão excluídos da MCL, antes de adicionar os demais candidatos (ROSEWARNE et al., 2015).

A derivação do candidato temporal do modo *Merge* é realizada observando-se o bloco co-localizado no quadro temporalmente mais próximo do quadro atual na lista de quadros de referência. A Figura 10 apresenta o bloco co-localizado  $Y$  e os dois candidatos  $C_0$  e  $C_1$ . Caso a PU na posição  $C_0$  não esteja disponível, seja codificada com predição intraquadros ou não pertença à CTU atual, a posição  $C_1$  é utilizada. Caso contrário, a posição  $C_0$  é utilizada para derivar o candidato temporal do *Merge* (ROSEWARNE et al., 2015).

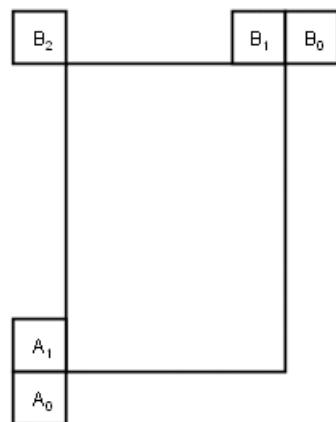


Figura 9: Posições de candidatos espaciais do *Merge* (ROSEWARNE et al., 2015)

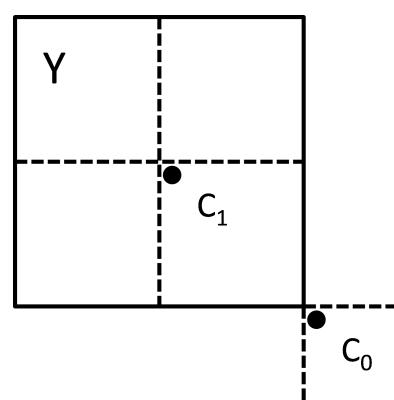


Figura 10: Possíveis posições para o candidato temporal do *Merge* em relação ao bloco co-localizado  $Y$  (ROSEWARNE et al., 2015)

Além dos candidatos espaciais e do candidato temporal, existem também mais dois tipos de candidatos *Merge*: candidato bi-preditivo e candidato de movimento zero. O primeiro, dado um conjunto de regras predefinidas, combina dois candidatos que já estejam na MCL, gerando um candidato *Merge* com bi-predição. Por fim, visando preencher a MCL até que esta atinja seu tamanho máximo, são adicionados na lista candidatos que não possuam deslocamento espacial em relação ao quadro de referência. O índice do quadro de referência utilizado começa em zero e é incrementado a cada novo candidato de movimento zero adicionado na MCL (ROSEWARNE et al., 2015).

Caso o modo *Merge* seja definido como melhor modo para uma dada PU, o codificador transmite uma *flag* indicando a seleção do modo *Merge* para este bloco, enviando juntamente o índice da MCL a qual o bloco codificado com *Merge* herdou os dados de movimento. Por fim, também são transmitidas as informações resultantes da codificação residual, geradas a partir da aplicação da MC utilizando os dados de movimento herdados pelo *Merge*.

### **2.2.5 Modo *Skip***

No HEVC, o modo *Skip* é tratado como um caso especial do modo *Merge*, sendo um dos principais responsáveis pela elevada taxa de compressão obtida no HEVC. Isto deve-se ao fato de que este modo é bem adequado para codificar imagens onde o erro de predição tende a ser muito baixo, descartando a necessidade de transmitir informação residual (SZE; BUDAGAVI; SULLIVAN, 2014).

No HEVC, quando uma CU é sinalizada utilizando o modo *Skip*, isto implica que estas três condições foram atendidas:

- esta CU contém apenas uma PU (particionamento 2Nx2N),
- o modo *Merge* foi utilizado para derivar os dados de movimento,
- o bitstream não contém informação residual.

Desta forma, as únicas informações transmitidas pelo codificador são: uma *flag* indicando que o modo *Skip* foi selecionado e o índice da MCL, a qual os dados de movimento foram herdados.

Resultados de simulação realizados em (SZE; BUDAGAVI; SULLIVAN, 2014) demonstram que é possível reduzir, em média, 7% do tamanho do vídeo codificado mantendo-se a mesma qualidade objetiva de imagem ao ativar os modos *Skip* e *Merge* no HEVC.

## 2.3 Considerações Finais do Capítulo

Esse capítulo apresentou os principais conceitos sobre a codificação de vídeos e uma breve introdução ao padrão HEVC. Foram apresentadas, de forma resumida, a estrutura de codificação do HEVC, bem como suas principais ferramentas de codificação. Estas informações serão importantes para a compreensão deste trabalho.

A revisão bibliográfica apresentada neste capítulo demonstrou que diversos modos e ferramentas de codificação foram integrados ao HEVC. Estas, somado às estruturas de particionamento de blocos flexíveis, contribuem para as altas taxas de compressão atingidas pelo padrão. Entretanto, esta eficiência de codificação é obtida ao custo de um elevado esforço computacional necessário para testar as diversas possibilidades de modos de predição, particionamento e demais ferramentas.

Dentre as ferramentas de codificação empregadas no HEVC, destaca-se o modo *Skip*. Este, por ser um caso especial do modo *Merge*, reduz a quantidade de informação lateral enviada ao decodificador. Somando-se a isto, este modo não transmite informações residuais do bloco codificado, portanto, infere-se que o *Skip* é um dos principais modos responsáveis pelas altas taxas de compressão obtidas no HEVC.

O próximo capítulo da dissertação apresentará os conceitos mais importantes sobre vídeos 3D e sobre o padrão 3D-HEVC, extensão do padrão HEVC para a codificação de vídeos 3D. Além disso, serão apresentadas as principais ferramentas do 3D-HEVC para a exploração de informações intervistas e codificação de mapas de profundidade. A apresentação dos conceitos básicos de vídeos digitais e a breve introdução ao HEVC descritas nas seções anteriores, juntamente com a revisão bibliográfica sobre o 3D-HEVC realizada no capítulo seguinte, possibilitarão estabelecer o cenário onde a heurística proposta nesta dissertação está inserida.

### **3 O PADRÃO 3D-HEVC**

Este capítulo apresenta o padrão de codificação de vídeos 3D-HEVC, que foi desenvolvido como uma extensão do padrão HEVC. Por este motivo, o capítulo anterior apresentou as estruturas de bloco, tal como as técnicas e ferramentas de codificação definidas pelo HEVC, que foram herdadas pelo 3D-HEVC. A revisão bibliográfica apresentada nesse capítulo foi realizada a partir do *Test Model 11* do 3D-HEVC (CHEN et al., 2015), do *Overview* das extensões multivista e 3D do HEVC (TECH et al., 2015) e também da descrição do 3D-HEVC(MULLER et al., 2013).

#### **3.1 Vídeos em três dimensões**

Vídeos em três dimensões (3D) possibilitam uma experiência visual mais elevada com a percepção de profundidade através do uso de *displays* especiais, os quais projetam a cena tridimensional com uma pequena diferença entre a imagem exibida ao olho esquerdo e direito do espectador.

Como exemplo de *displays* especiais para este tipo de aplicação, pode-se citar os *displays* estereoscópicos, os quais basicamente apresentam as duas vistas que foram originariamente capturadas por uma câmera estereoscópica. Neste cenário, faz-se necessário que os espectadores utilizem óculos que filtrem a imagem a ser exibida para seu olho direito e esquerdo(TECH et al., 2015). Outro *display* especial para reproduzir vídeos 3D são os *displays* multivistas (auto estereoscópicos), os quais não apresentam apenas um par de imagens, mas um conjunto bem mais numeroso de vistas (entre 20 e 50) (MULLER et al., 2013). Embora este sistema ainda faça com que os espectadores percebam diferentes imagens em diferentes angulações no olho esquerdo e direito, um *display* multivista possibilita que os espectadores assistam diferentes pares de imagens em diferentes posições. Assim, não apenas elevando a percepção visual da imagem ao espectador, mas permitindo a percepção da profundidade do vídeo sem o uso de óculos 3D (TECH et al., 2015).

Os primeiros formatos de vídeos multivista propostos continham apenas informações de vídeo capturadas pelas câmeras, como o formato convencional de

vídeo estéreo (*Conventional Stereo Video - CSV*) e o formato de vídeo multivista (*Multiview Video - MVV*) (MULLER et al., 2013). Visando codificar eficientemente este tipo de mídia, desenvolveu-se o padrão Multiview Video Coding (MVC) (JVT, 2009), o qual fora desenvolvido como uma extensão do padrão H.264/AVC. Devido à exploração de informação entre diferentes vistas, o MVC provê uma maior eficiência na compressão do vídeo quando comparado à codificação de cada vista de forma independente. Entretanto, para manter uma determinada qualidade de imagem, o *bitstream* do vídeo codificado tende a crescer linearmente com o número de vistas codificadas (MULLER et al., 2013). Desta forma, a transmissão de um vídeo codificado com o MVC, para um sistema que utilize diversas vistas torna-se inviável.

Visando contornar este problema, introduziu-se o formato de vídeos multivista mais profundidade (*Multiview plus Depth - MVD*), onde informações da distância dos objetos da cena até a câmera (mapas de profundidade) são capturadas para cada vista (textura). Desta forma, padrões de codificação de vídeo que utilizem este formato precisam codificar apenas um conjunto reduzido de vistas, juntamente com seus respectivos mapas de profundidade. As demais vistas podem ser geradas pelo sistema que recebe o vídeo codificado a partir de técnicas de síntese de vistas, como a Renderização Baseada em Imagem de Profundidade (*Depth Image Based Rendering - DIBR*) (KAUFF et al., 2007).

Diferentemente do MVC, o 3D-HEVC apresenta suporte a codificação a vídeos em formato MVD, elevando ainda mais a eficiência de codificação do HEVC em relação ao H.264/AVC neste cenário (TECH et al., 2015).

A Figura 11 exemplifica um sistema utilizando duas vistas com seus respectivos mapas de profundidade (Mapa Prof.). É importante observar que, uma vez que a cena é capturada por mais de uma câmera simultaneamente, percebe-se o surgimento da correlação intervistas, permitindo a exploração das redundâncias observadas neste cenário pelo codificador. Assim, o codificador 3D-HEVC, além de efetuar a predição intraquadro e interquadros, também expande sua eficiência de codificação através da predição intervistas. Destaca-se que estas três formas de predição também são aplicáveis aos mapas de profundidade. Por fim, a Figura 11 também ilustra um exemplo de vista sintetizada intermediária 0,5, a qual é gerada utilizando as vistas 0 e 1 e seus respectivos mapas de profundidade.

### **3.2 Sistema de Codificação do 3D-HEVC**

O conceito básico de um sistema baseado no 3D-HEVC, com suporte ao formato MVD, é ilustrado na Figura 12. Em linhas gerais, a entrada do codificador 3D-HEVC consiste em diversos quadros de textura e de mapa de profundidade de diversas vistas, sendo uma destas vistas (chamada de vista independente), codificada sem utilizar

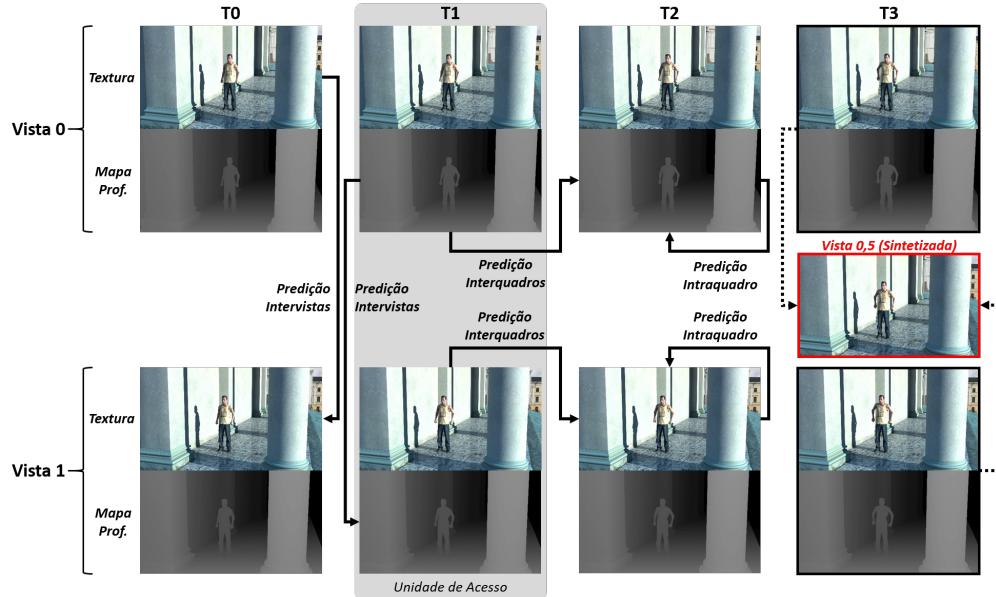


Figura 11: Exemplo de um sistema com duas vidas e seus respectivos mapas de profundidade

informações intervistas. A partir da codificação, o 3D-HEVC gerará um *bitstream* que representa o vídeo de entrada somado às informações de profundidade em um formato codificado. Caso este *bitstream* seja decodificado por um decodificador 3D, o vídeo somado ao mapa de profundidade será reconstruído com uma dada fidelidade.

Para a exibição do vídeo 3D em um *display* auto estereoscópico, vistas intermediárias podem ser geradas através do DIBR utilizando as imagens de textura e mapas de profundidade do vídeo decodificado. No caso em que o decodificador esteja conectado em um *display* estereoscópico, o sintetizador de vidas também pode gerar o par de vidas, caso estas não estejam presentes no *bitstream*. Neste caso, é possível ajustar as vidas do vídeo renderizado para a geometria estéreo das condições de visualização. Uma das vidas decodificadas ou uma vista sintetizada arbitrária também pode ser utilizada para ser exibida em um *display* 2D convencional (CHEN et al., 2015).

O *bitstream* gerado pelo 3D-HEVC é construído de uma forma que o sub-*bitstream* que representa a codificação da vida independente pode ser facilmente extraído. Assim, é possível descartar os pacotes que contêm informações das vidas dependentes e dos mapas de profundidade, possibilitando a decodificação do *bitstream* restante por um decodificador HEVC convencional (CHEN et al., 2015).

### 3.3 Ordem de Codificação do 3D-HEVC

Uma vez que o HEVC codifica vídeos que contenham apenas uma vida e informações de textura, cada quadro deste representa uma cena de um determinado

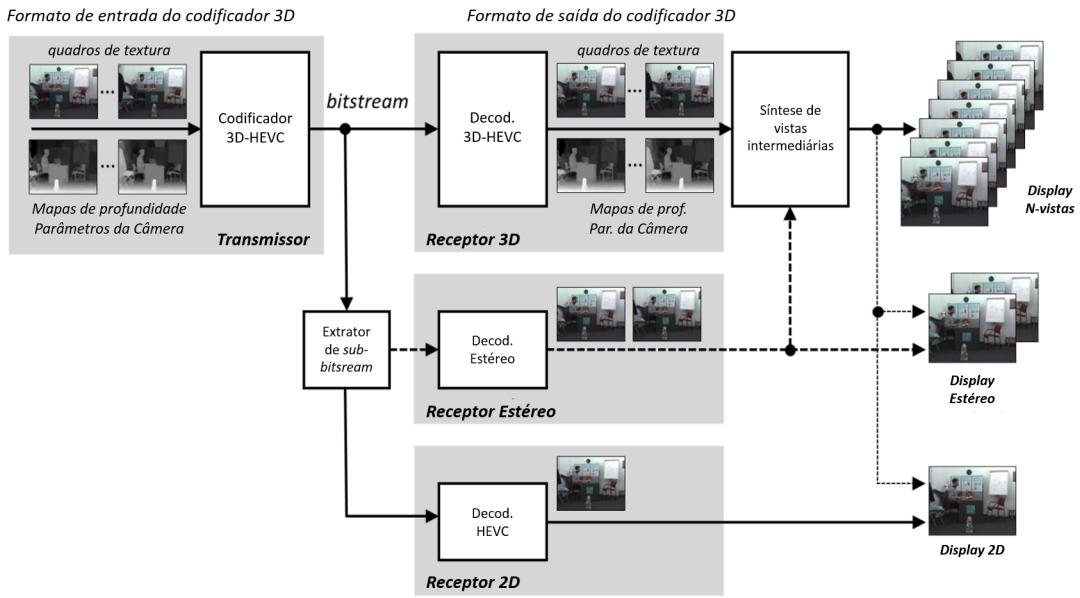


Figura 12: Visão geral da estrutura e formato de dados para transmissão de vídeos 3D. Adaptada de (CHEN et al., 2015)

instante de tempo, diferentemente do que ocorre na representação de um vídeo em formato MVD, MVV ou até mesmo CSV.

Ao codificar um vídeo no formato MVD, o 3D-HEVC define cada conjunto de imagens que pertencem ao mesmo instante de tempo como unidades de acesso (Access Units - AU). Cada AU engloba todos os quadros de textura, e seus respectivos mapas de profundidade, que pertencem ao mesmo instante de tempo. A Figura 13 exemplifica como são formadas as AUs, onde *vistaID* define a que vista cada quadro de textura (*text.*) e mapa de profundidade (*prof.*) pertencem. O sistema apresentado na Figura 13 contém  $N$  vistas, numeradas de zero até  $N - 1$ .

Os quadros de textura da vista número zero (*vistaID* 0) de cada AU são sempre codificadas com um codificador HEVC sem alterações, assim, a vista zero é denominada vista independente. As demais vistas são processadas pelo codificador capaz de observar redundâncias intervistas, denominadas então de vistas dependentes. O codificador de mapas de profundidade, embora apresente algumas modificações em relação ao codificador de quadros de textura, segue o mesmo sistema, onde o mapa de profundidade da vista independente não utiliza informações intervistas, ao passo que os demais utilizam estas informações (CHEN et al., 2015). A Figura 14 apresenta a estrutura básica do codificador de vídeos 3D-HEVC. O bloco azul ilustra o codificador HEVC sem alterações, que é capaz de gerar um *bitstream* compatível com o decodificador 2D. As linhas vermelhas apresentam as informações intervistas e inter-componentes (textura → mapa de profundidade) transmitidas entre os “codificadores internos” (textura, mapa de profundidade, independente e dependente). Por fim, as linhas vermelhas pontilhadas representam as informações inter-componentes trans-

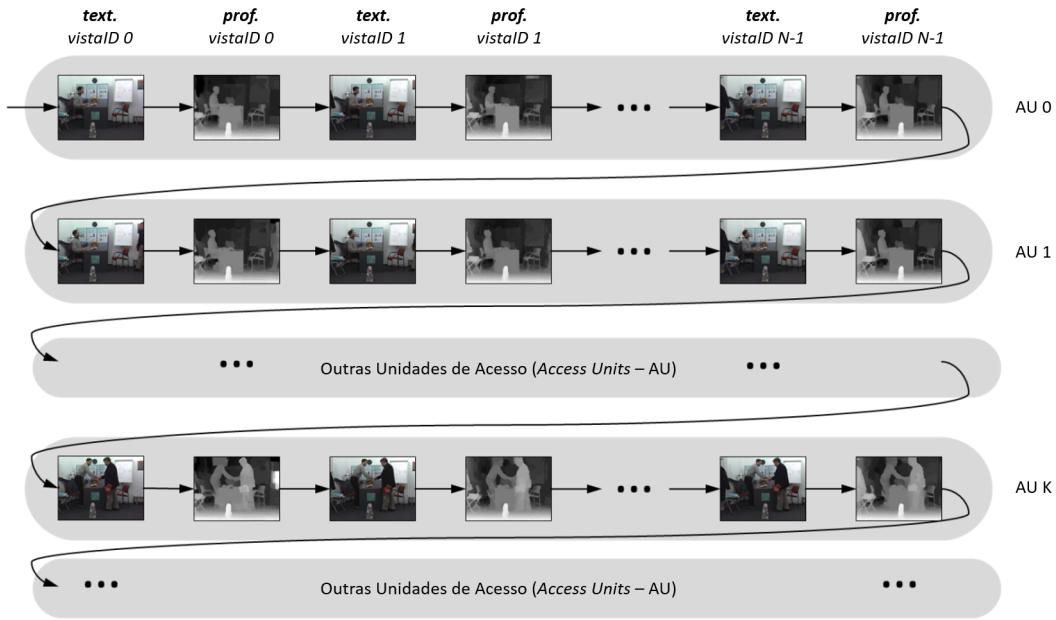


Figura 13: Estrutura de Unidades de Acesso do 3D-HEVC. Adaptada de (CHEN et al., 2015)

mitidas do codificador de mapa de profundidade de uma dada vista para o codificador de textura de outra vista.

### 3.4 Codificação das Vistas Dependentes no 3D-HEVC

Uma vez que a vista independente, também referida como vista base, é processada por um codificador HEVC original, esta seção discutirá o processo de codificação de vistas dependentes do 3D-HEVC. É importante ressaltar que os mesmos conceitos e ferramentas utilizadas na codificação da vista base também são empregados na codificação das vistas dependentes, além das ferramentas adicionais apresentadas nesta seção.

#### 3.4.1 Predição Intervistas

A predição intervistas é muito semelhante a predição interquadros do HEVC, sendo basicamente diferenciadas pela fonte do quadro a ser utilizado como referência. Enquanto a predição interquadros obtém a MCP utilizando como referência quadros previamente codificados que pertencem a outros instantes de tempo, a predição intervistas visa obter a predição compensada pela disparidade (*Disparity-Compensated Prediction* - DCP) utilizando quadros de outras vistas previamente codificadas, entretanto pertencentes ao mesmo instante de tempo (CHEN et al., 2015).

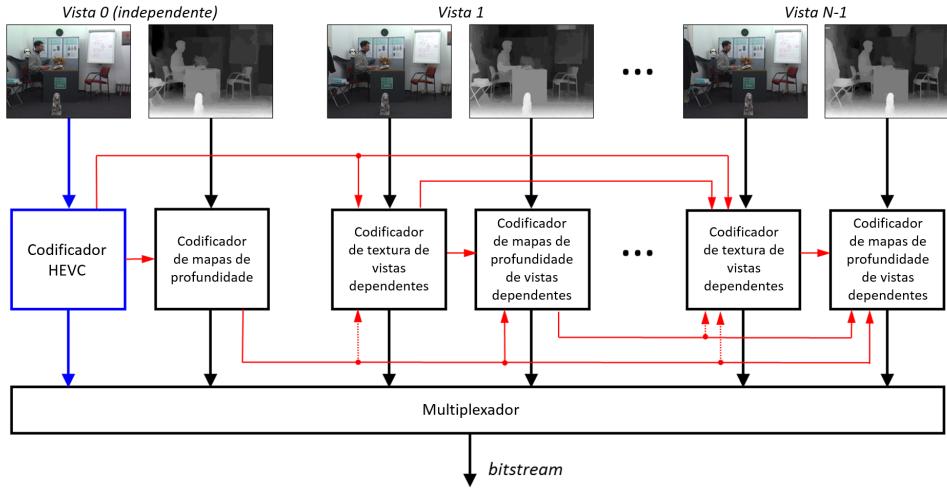


Figura 14: Estrutura básica do codificador 3D-HEVC com informações intervista e inter-componente. Adaptada de (CHEN et al., 2015)

### 3.4.2 Obtenção do Vetor de Disparidade

O vetor de disparidade é utilizado com o objetivo de localizar o bloco correspondente da PU/CU atual em outras vidas no mesmo instante de tempo. Além da predição intervistas, o vetor de disparidade também é útil ao 3D-HEVC no emprego de outras técnicas de codificação, tais como a Predição de Movimento Intervistas (*Interview Motion Prediction* - IMP), Predição Avançada de Resíduos (*Advanced Residual Prediction* - ARP), Compensação de Iluminação (*Illumination Compensation* - IC) e Predição Baseada em Síntese de Vistas (*View Synthesis Prediction* - VSP) (CHEN et al., 2015).

A obtenção dos vetores de disparidade do bloco atual no 3D-HEVC é realizada derivando os vetores de disparidade de blocos vizinhos (*Neighboring Block Disparity Vector* - NBDV) espaciais e temporais previamente codificados. Primeiramente, os blocos DCP temporalmente vizinhos são avaliados, seguido pela avaliação dos blocos vizinhos espaciais. Por fim, blocos vizinhos codificados com predição MCP são buscados e avaliados. Uma vez que um vetor de disparidade é encontrado, todo o processo de obtenção deste encerra.

Até dois quadros de referência são utilizados a partir da vista atual para a obtenção dos blocos vizinhos temporais do NBDV. O primeiro quadro candidato é o quadro co-localizado, o qual é previamente definido pelo codificador. O segundo quadro é obtido nas listas de quadros de referência e adicionado dentro da lista de candidatos através de um esquema de acesso aleatório) (CHEN et al., 2015). Para cada quadro de referência candidato, apenas um bloco temporalmente vizinho é buscado. Apenas dois vizinhos são avaliados na obtenção do NBDV a partir de blocos vizinhos espaciais, sendo estes o bloco inferior esquerdo e o bloco superior direito da CU atual, conforme apresentado na Figura 15. O processo de avaliação dos blocos espaciais começa com

o bloco  $A_1$  e termina com o bloco  $B_1$ . A Figura 15 também apresenta a posição dos blocos temporais buscados na lista de quadros candidatos, indicado por  $T_0$ .

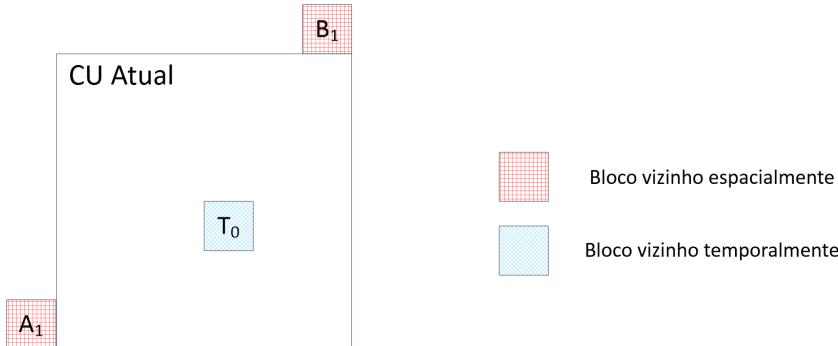


Figura 15: Localização dos blocos vizinhos temporais e espaciais do NBDV. Adaptada de (CHEN et al., 2015)

A obtenção do vetor NBDV também pode ser realizada a partir de blocos vizinhos codificados com predição interquadros, caso tenham seu movimento predito através da predição IMP (apresentada a seguir). O vetor de disparidade utilizado para a predição IMP representa a correspondência de movimento entre o quadro atual e o quadro de referência intervistas. O tipo de vetor de movimento gerado pela IMP é denominado de vetor de movimento predito por entrevistas (*Interview Predicted Motion Vector* - IvpMv) e os blocos codificados com este tipo de predição são denominados como predição compensada por movimento utilizando vetor de disparidade (*Disparity Vector* - *Motion Compensated Prediction* - DvMCP).

Resume-se que o processo completo de obtenção do vetor NBDV primeiramente verifica se há algum bloco vizinho temporal ou espacial que tenha sido codificado com DCP. Caso nenhum seja encontrado, o processo de derivação do vetor de disparidade faz a verificação de blocos vizinhos espaciais compensados por blocos do tipo DvMCP nas posições  $A_1$  e  $B_1$  da Figura 15, seguindo esta ordem (CHEN et al., 2015).

### 3.4.3 Refinamento do Vetor de Disparidade

Durante a codificação de um quadro de textura de uma dada vista dependente, o mapa de profundidade reconstruído da vista base já está disponível. Desta forma, é possível obter um vetor de disparidade mais preciso que o gerado pelo NBDV, utilizando as informações da distância entre os componentes da cena até a câmera da vista de base. Assim, obtém-se o vetor de disparidade baseado em blocos vizinhos orientado pela profundidade (*Depth-oriented Neighboring Block Disparity Vector* - DoNBDV) (CHEN et al., 2015).

### 3.4.4 Modo Merge

O modo *Merge* do 3D-HEVC sofre algumas alterações quando comparado ao HEVC, principalmente na construção da MCL. Isto se deve ao fato que a correlação entre as vistas pode ser explorada, predizendo os vetores de movimento das vistas dependentes, levando em conta que duas vistas diferentes representam a mesma cena capturada em ângulos diferentes. Assim, há uma grande probabilidade que os blocos co-localizados em quadros intervistas previamente codificados possuam as mesmas informações de movimento (TECH et al., 2015).

O processo de obtenção dos candidatos *Merge* é basicamente dividido em dois passos. O primeiro consiste repetir os passos utilizados para a construção da MCL conforme especificado pelo HEVC. No segundo passo, a extensão MCL para o 3D-HEVC é feita a partir da primeira lista, além de adicionar novos candidatos de forma intercalada aos candidatos da lista inicial. Os candidatos adicionais da MCL estendida são: os candidatos intervistas, o candidato da predição de síntese de vista e o candidato de disparidade (TECH et al., 2015).

É importante ressaltar que o processo de geração da MCL para blocos de textura é diferente do processo para os blocos de mapas de profundidade. Além disso, destaca-se que o processo discutido até então é referente ao processo da geração dos candidatos *Merge* para blocos pertencentes a quadros de textura. A geração da MCL estendida na codificação dos mapas de profundidade é bem mais simples, utilizando apenas candidatos provenientes da textura e da predição intervistas.

### 3.4.5 Predição de Movimento Intervistas

A predição de movimento intervistas (*Interview Motion Prediction* - IMP) assume que há uma estimativa correta do mapa de profundidade para a imagem de textura que está sendo codificada. Também é importante destacar que a IMP é aplicada pelo modo *Merge* na geração dos candidatos da MCL.

O funcionamento da IMP é apresentado na Figura 16, onde o bloco que está sendo codificado está na localização  $x$ . O valor máximo  $d$  dentro do bloco co-localizado na profundidade é convertido em um vetor de disparidade. Este vetor de disparidade é utilizado para encontrar o bloco co-localizado na vista de referência na posição  $xR$ . Se o bloco de referência foi codificado utilizando MCP, os parâmetros de movimento associados podem ser utilizados como parâmetros de movimento candidatos para o bloco atual na vista atual. O vetor de disparidade também pode ser usado diretamente como um candidato de vetor de disparidade para a DCP (CHEN et al., 2015).

### 3.4.6 Predição Avançada de Resíduos

A predição avançada de resíduos (*Advanced Residual Prediction* - ARP) é uma ferramenta de codificação que explora a correlação residual entre vistas pertencentes

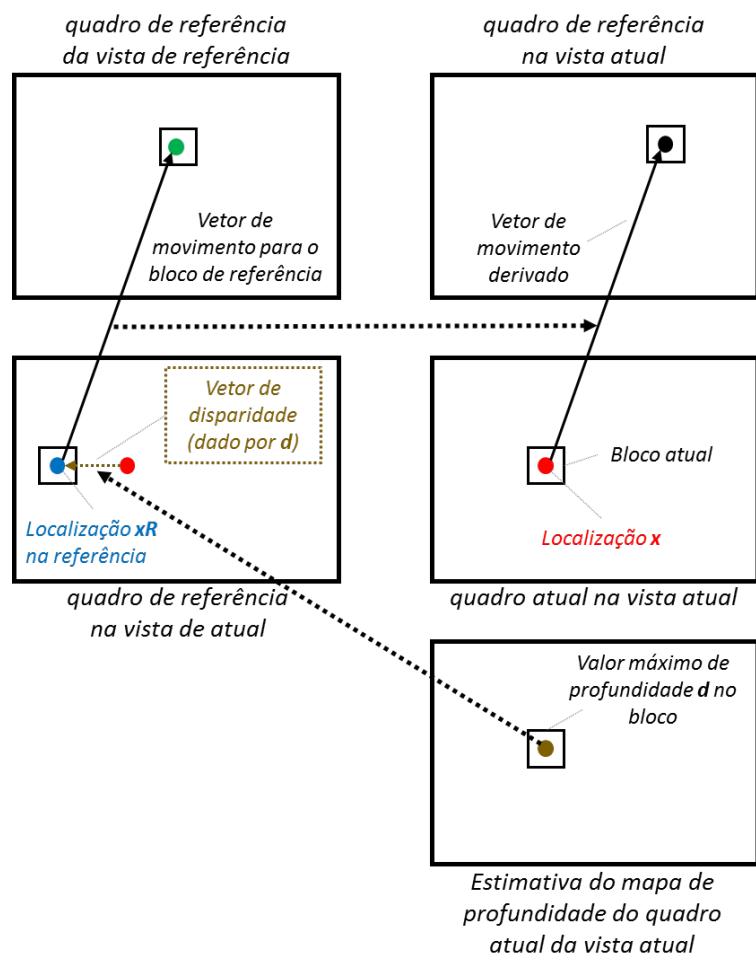


Figura 16: Processo de obtenção dos parâmetros de movimento para um bloco no quadro atual baseado em parâmetros de movimento da vista de referência já codificada e uma estimativa do mapa de profundidade para o quadro atual. Adaptada de (CHEN et al., 2015)

à mesma Unidade de Acesso, aumentando ainda mais a eficiência de codificação das vistas dependentes. Um preditor residual é produzido alinhando as informações de movimento da vista atual predita por MCP na vista de referência. Somando-se a isto, fatores de ponderação são utilizados visando compensar a diferença de qualidade entre as diferentes vistas.

Quando a ferramenta ARP é habilitada para um determinado bloco, somente a diferença entre o resíduo do bloco atual e do resíduo predito é considerada no processo de codificação. Cabe ressaltar que esta técnica somente é aplicada em PUs 2Nx2N (64x64, 32x32, 16x16 e 8x8) (CHEN et al., 2015).

#### **3.4.7 Compensação de Iluminação**

A compensação de iluminação (*Illumination Compensation - IC*) é utilizada no 3D-HEVC visando adaptar as amostras de luminância e crominância dos blocos intervisitas preditos para a iluminação da vista atual. Desta forma, reduz-se o erro gerado pela predição. O processo é basicamente realizado através de um modelo linear, utilizando uma solução de erro quadrático mínimo (CHEN et al., 2015).

#### **3.4.8 Predição Baseada em Síntese de Vistas**

A predição baseada em síntese de vistas (*View Synthesis Prediction - VSP*) provê um preditor para cada PU utilizando informações de profundidade. A imagem predita é renderizada deslocando as amostras da vista de base utilizando as informações de disparidade obtidas através do mapa de profundidade correspondente da vista de base (TECH et al., 2015).

#### **3.4.9 Particionamento de Bloco Baseado em Profundidade**

O modo de particionamento de bloco baseado em profundidade (*Depth-Based Block Partitioning - DBBP*) partitiona o bloco de textura usando uma máscara de segmentação binária (*Binary Segmentation Mask - BSM*) (CHEN et al., 2015), que é obtida através do mapa de profundidade correspondente. Cada uma das duas partições tem seu movimento predito de forma isolada, sendo então reincorporadas novamente na BSM. É importante salientar que o DBBP é habilitado apenas para tamanhos de blocos maiores que 8x8.

### **3.5 Codificação de Mapas de Profundidade no 3D-HEVC**

No processo de codificação dos mapas de profundidade, são utilizados os mesmos conceitos básicos de predição intraquadro, MCP, DCP, etc., utilizados na textura. Entretanto, algumas ferramentas foram modificadas, algumas desabilitadas e outras até mesmo adicionadas neste cenário visando codificar eficientemente os ma-

pas de profundidade, dadas as suas características específicas. Como exemplo, as ferramentas IMP, ARP e VSP, apresentadas previamente, estão desabilitadas no processo de codificação de mapas de profundidade. Isto ocorre, pois, a ordem de codificação (textura e então mapa de profundidade) possibilita obter os parâmetros de movimento baseado no bloco correspondente de textura, simplificando o processo de codificação (CHEN et al., 2015).

Visto que os mapas de profundidade são basicamente formados por arestas bem definidas e regiões homogêneas, diferentemente das vistas de textura, o uso do filtro de interpolação da predição interquadros prejudicaria a qualidade dos mapas ao atenuar as bordas de objeto, acarretando no comprometimento do processo de síntese de vistas. Por este motivo, a predição interquadros é modificada na codificação de mapas de profundidade, desabilitando a estimativa de movimento fracionária, a qual utiliza o filtro de interpolação (SULLIVAN et al., 2012).

Por fim, visto que os filtros *Deblocking Filter* (DF) e *Sample Adaptive Offset* (SAO) (SULLIVAN et al., 2012) do HEVC foram implementados focando nas características de imagem dos quadros de textura, estes tendem a não ser eficientes quando aplicados a mapas de profundidade. Desta forma, assim como as ferramentas IMP, ARP e VSP, os filtros foram desabilitados no processo de codificação de mapas de profundidade no 3D-HEVC.

A seguir, serão apresentadas as ferramentas criadas especificamente para o codificador de mapas de profundidade do 3D-HEVC.

### **3.5.1 Modos Intra para Modelagem de Profundidade**

Dado que os mapas de profundidade são principalmente caracterizados por arestas bem definidas e por extensas áreas com valores constantes - ou praticamente constantes - de amostras, a codificação intraquadro convencional estipulada pelo HEVC tende a ser menos eficiente neste cenário. Isto se deve ao fato de que os 35 modos da predição intra utilizados no HEVC são eficientes em regiões com valores de amostras praticamente constantes, gerando artefatos significativos em regiões com arestas bem definidas, assim, comprometendo o processo de geração de vistas sintetizadas.

Para uma melhor representação das arestas presentes nos mapas de profundidade, dois novos modos de predição intraquadro, denominados modos para modelagem de profundidade (*Depth Modelling Modes* - DMM) (CHEN et al., 2015), foram desenvolvidos. Em cada um destes modos, o bloco de profundidade é estimado através de um modelo que divide a área do bloco em duas regiões não retangulares, onde cada uma destas é representada por uma constante. Assim, duas informações são utilizadas neste processo: a) a especificação de qual partição cada amostra do bloco pertence; e b) a constante de cada uma das regiões, denominado valor constante da

partição (*Constant Partition Value* - CPV).

Os dois modos DMMs definem que tipo de partição será utilizado para estimar os valores das amostras dos blocos, podendo ser baseados em retas (*Wedgelets*) ou em contorno (*Countour*). Assim, estes dois tipos de partição se diferem na forma como a segmentação do bloco é realizada. No caso da partição do tipo *Wedgelet*, ilustrada na Figura 17, as duas regiões ( $P_1$  e  $P_2$ ) são separadas por uma linha reta, cujo ponto inicial desta é representado por  $S$  e ponto final é representado por  $E$ . Naturalmente, os pontos  $S$  e  $E$  não podem pertencer à mesma borda do bloco.

A Figura 17 (a) apresenta a linha de separação das regiões no espaço contínuo, no qual a linha pode ser descrita por uma equação de reta utilizando os pontos que definem o fim do bloco no eixo  $x$  ( $x_B$ ) e no eixo  $y$  ( $y_B$ ) juntamente com os pontos que definem o início e o fim da reta ( $x_S$  e  $x_E$  respectivamente). A Figura 17 (b) apresenta dados semelhantes aos da Figura 17 (a), entretanto, representadas no espaço discreto. Destaca-se que as regiões  $P_1$  e  $P_2$  não são as mesmas nas Figuras 17 (a) e (b). Isto ocorre, pois, no espaço discretizado nem todas as amostras estão completamente em uma das regiões, fazendo com que algumas amostras de uma determinada região avancem a linha divisória. Por fim, a Figura 17 (c) apresenta como as regiões  $P_1$  e  $P_2$  são representadas no processo de codificação.

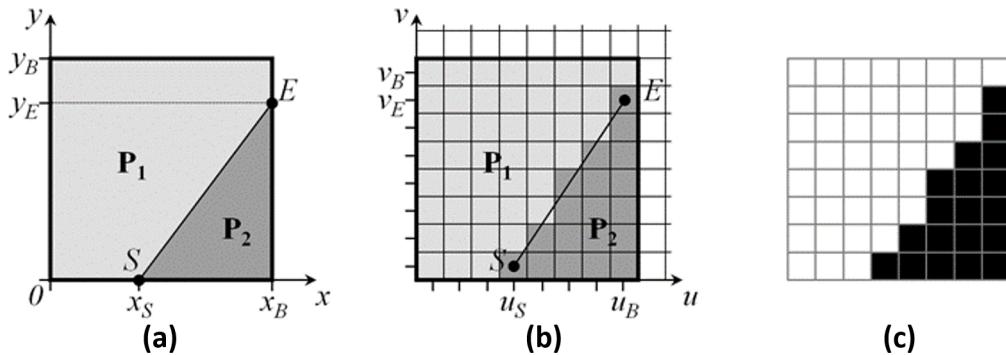


Figura 17: DMM *Wedgelet* de um bloco de profundidade: (a) espaço contínuo, (b) espaço discreto, (c) padrão de partição correspondente (CHEN et al., 2015)

Diferentemente das *Wedgelets*, a linha de separação entre duas regiões do modo DMM *Countour* pode não ser facilmente descrita através de uma função geométrica, uma vez que as regiões  $P_1$  e  $P_2$  podem ter formas arbitrárias, além de conter diversas partes. Um exemplo de partição do DMM *Countour* é apresentado na Figura 18, onde (a) apresenta a partição no espaço contínuo, (b) apresenta a partição no espaço discreto e por fim (c) apresenta o padrão de partição correspondente.

Além do particionamento, o DMM utiliza outra informação no processo de codificação, que é o CPV de cada uma das regiões. Visto que o particionamento visa separar duas regiões que contenham valores praticamente constantes de amostras, o

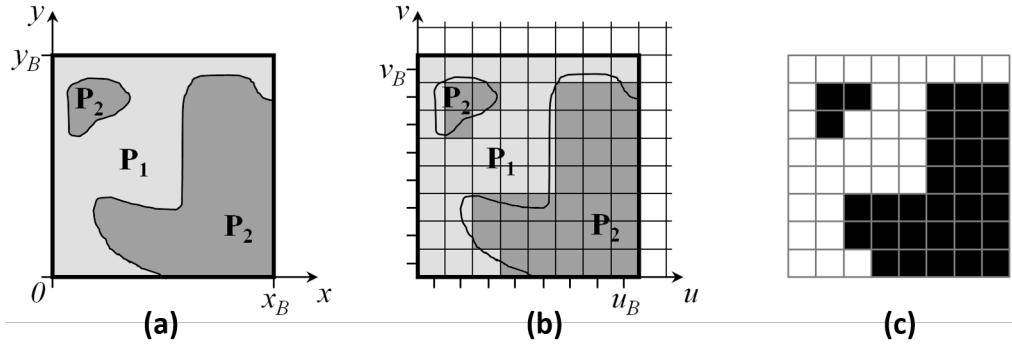


Figura 18: DMM *Contour* de um bloco de profundidade: (a) espaço contínuo, (b) espaço discreto, (c) padrão de partição correspondente (CHEN et al., 2015)

CPV de cada uma das regiões é obtido utilizando-se o valor médio das amostras de profundidade original pertencentes à região em questão (CHEN et al., 2015).

### 3.5.2 Codificação Segment-wise DC

A abordagem de codificação *Segment-wise DC* (SDC) é um método alternativo à codificação residual, onde os resíduos não são processados pelos módulos convencionais da transformada e quantização (CHEN et al., 2015).

A codificação SDC pode ser aplicada na informação residual gerada pela predição intraquadro dos mapas de profundidade. No caso em que o resíduo tenha sido gerado por um dos modos de predição intraquadro convencionais do HEVC, a PU inteira é considerada como um único segmento, sendo então gerado para ela apenas um valor residual. No caso em que o resíduo tenha sido gerado a partir da aplicação de um modo DMM, a PU é dividida em dois segmentos (conforme o particionamento do DMM), desta forma, dois valores residuais são gerados.

Antes de processar os dados, o SDC mapeia os resíduos para valores que estão no mapa de profundidade, utilizando uma tabela de pesquisa de profundidade (*Depth Lookup Table - DLT*) (CHEN et al., 2015). Desta forma, os valores residuais podem ser codificados apenas sinalizando o índice na DLT, reduzindo a largura de bits necessária para representar o resíduo. Este mapeamento é transmitido para o decodificador, o qual converte o índice novamente em um valor de profundidade. A DLT é previamente construída pelo codificador, o qual observa um número pré-definido de mapas de profundidade originais, gerando então a DLT. Esta é transmitida para o decodificador juntamente com as informações de imagem.

### 3.5.3 Herança de Parâmetros de Movimento

Visto que os quadros de textura e de mapas de profundidade de uma mesma vista apresentam a mesma cena, entretanto com abordagens diferentes, o movimento de ambos tende a apresentar grande similaridade (TECH et al., 2015). Desta forma,

visando aumentar a eficiência do processo de codificação dos mapas de profundidade, o modo de herança de parâmetros de movimento (*Motion Parameters Inheritance - MPI*) é um dos candidatos da MCL do modo *Merge*. O 3D-HEVC utiliza a sintaxe do modo *Merge*, visto que é possível efetuar a sinalização eficiente no caso onde não exista resíduos (Modo *Skip*) (TECH et al., 2015).

Uma vez que os vetores de movimento da textura possuem precisão de até 1/4 de pixel ao passo que os vetores de movimento dos mapas de profundidade são inteiros, os vetores herdados são quantizados para o valor inteiro mais próximo (CHEN et al., 2015).

### **3.5.4 Codificação *Segment-wise DC* no Modo *Inter***

A codificação *Segment-wise DC* no Modo *Inter* (*Inter-mode segment-wise DC coding - ISDC*) estende a ideia do SDC para a codificação residual a partir da aplicação de modos interquadros. Desta forma, as etapas de Transformada e Quantização não são utilizadas juntamente com o ISDC. Visando evitar uma sobreposição de modos, o ISDC não é aplicado em CUs que tenham sido codificadas com *Skip* (CHEN et al., 2015). Desta forma, o ISDC somente é aplicado quando o resíduo DC, dentro de uma dada PU, é diferente de zero. O resíduo DC de uma PU é calculado através da média das diferenças entre os valores das amostras originais e os valores de amostras preditas de todas as amostras com a PU. O modo ISDC, assim como o *Skip*, somente é aplicado em PUs de partição 2Nx2N (CHEN et al., 2015).

### **3.5.5 Predição de *Quadtree* de Profundidade**

A predição de *Quadtree* de profundidade (*Depth Quadtree Prediction - DQP*) efetua a predição dos modos de particionamento da profundidade com base nas informações da textura. Salienta-se que a DQP é aplicada apenas em quadros que não pertençam à unidades de acesso *Random Access Point* (RAP). Uma unidade de acesso é RAP quando a vista de base (textura e profundidade) é codificada apenas com predição intraquadro (MULLER et al., 2013). Outra predição efetuada pela DQP é que uma CU de profundidade jamais pode ser menor que a CU correspondente de textura, realizando uma limitação na *quadtree* (*Quadtree Limitation - QTL*) de profundidade. A predição dos particionamentos de profundidade com base na textura é apresentada na Figura 19.

### **3.5.6 Modo *Skip Intra* de Profundidade**

Uma vez que os mapas de profundidade contêm regiões com valores bem homogêneos das amostras e, que determinadas perdas de qualidade nestas regiões não apresentam grande impacto na qualidade das vistas sintetizadas, o conceito de modo *Skip Intra* de profundidade (*Depth Intra Skip - DIS*) foi inserido no 3D-HEVC.

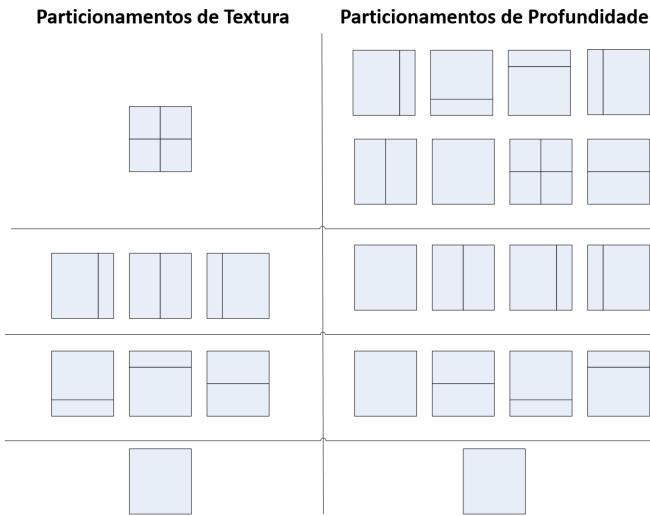


Figura 19: Particionamento da textura e os possíveis particionamentos de profundidade. Adaptada de (CHEN et al., 2015)

Desta forma, a codificação residual nestas áreas não é sinalizada, assim como ocorre no *Skip* convencional.

O modo DIS pode ser empregado de duas formas. Na primeira - ilustrada na Figura 20 -, é aplicado o modo intraquadro angular do HEVC convencional com direção (a) vertical ou (b) horizontal, sem armazenar informações residuais. Desta forma, o bloco predito é igual ao bloco reconstruído. A segunda forma do modo DIS utiliza um valor de profundidade único. O valor predito é escolhido a partir das duas amostras vizinhas espaciais reconstruídas da CU atual, as amostras  $A_{n/2}$  e  $B_{n/2}$  na Figura 21. O valor de uma amostra vizinha da CU atual é diretamente usado como valor de pixel do sinal de predição. Se estas amostras não estiverem disponíveis, o valor médio da faixa valores de profundidade é usado para toda CU. Um índice é utilizado para identificar o modo de predição da lista de candidatos da CU codificada com DIS (CHEN et al., 2015).

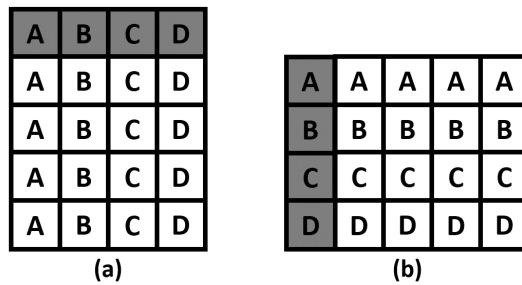


Figura 20: Modos de predição intraquadro (a) vertical e (b) horizontal utilizados pela ferramenta DIS

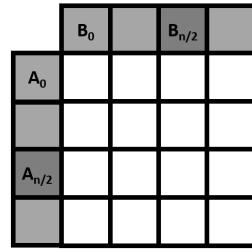


Figura 21: Amostras vizinhas espaciais reconstruídas utilizadas pela ferramenta DIS

### 3.6 Considerações Finais do Capítulo

Esse capítulo apresentou, de forma resumida, os principais conceitos sobre a codificação de vídeos 3D e sobre o padrão 3D-HEVC, que é uma extensão no padrão HEVC. As principais ferramentas adicionadas no 3D-HEVC em relação ao HEVC também foram resumidamente apresentadas. Assim, foi possível estabelecer o cenário onde este trabalho está inserido. Em especial, foi possível apresentar o grau de complexidade das ferramentas do 3D-HEVC e demonstrar relevância do desenvolvimento de soluções eficientes para a redução de complexidade deste padrão.

Dentre as diversas ferramentas propostas no 3D-HEVC, o modo DIS destaca-se pelo fato de, assim como o modo Skip convencional, não transmitir informações residuais. Desta forma, infere-se que blocos codificados com este modo utilizarão poucos bits em sua representação, atingindo elevados ganhos de compressão.

O próximo capítulo apresenta a avaliação realizada sobre o 3D-HEVC, desenvolvida para verificar a contribuição dos diversos modos de codificação definidos pelo 3D-HEVC tanto na eficiência de codificação, como no tempo de codificação. Através desta análise torna-se mais clara a contribuição principal deste trabalho.

## 4 ANÁLISES SOBRE A CODIFICAÇÃO DOS MAPAS PROFUNDIDADE NO 3D-HEVC

Este capítulo apresenta as análises realizadas sobre a codificação dos mapas de profundidade no 3D-HEVC, as quais visam avaliar as decisões adotadas pelo codificador a fim de auxiliar no desenvolvimento de heurísticas para redução da complexidade com os menores impactos na eficiência de codificação.

Visando facilitar a descrição das análises apresentadas ao longo deste capítulo, a Figura 22 ilustra os modos de particionamento de CU em PUs, conforme demonstrado na Figura 3, entretanto, distinguindo PUs retangulares e assimétricas. Embora um quadrado seja considerado um caso especial de um retângulo, onde todos os lados têm o mesmo comprimento, este trabalho considera que os modos de particionamento quadráticos não são retangulares, portanto, modos quadráticos são diferentes de modos retangulares.

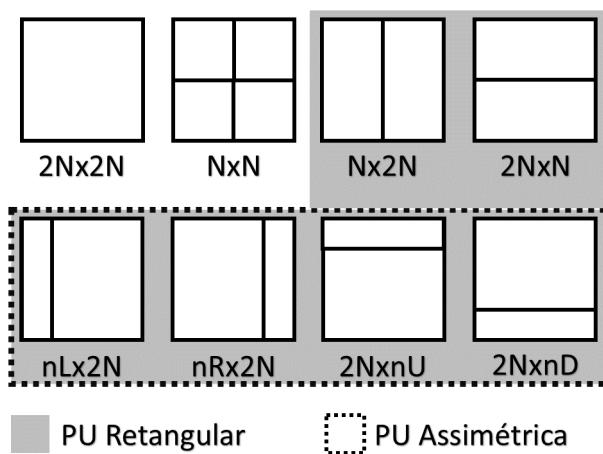


Figura 22: Modos de particionamento de CUs em PUs classificadas de acordo com seu formato

Como é possível perceber, o conjunto das PUs assimétricas está contido dentro do conjunto das PUs retangulares. Além disso, ressalta-se que os modos de particionamento que não pertencem ao conjunto de PUs retangulares são considerados

quadráticos.

Dado que este trabalho aborda uma estratégia de redução de complexidade para o codificador de mapas de profundidade do 3D-HEVC, este capítulo apresenta, inicialmente, uma análise da distribuição de modos no codificador de mapas de profundidade do 3D-HEVC. Assim, é possível detectar quais são os modos mais usados e, portanto, mais relevantes. A seguir, é realizada uma segunda análise, desta vez relacionando a taxa de bits, a qualidade de imagem e a redução de complexidade (*Rate-Distortion-Complexity* - RDC) ao desabilitar modos de particionamento pouco utilizados no processo de codificação. Estes resultados serão importantes para a compreensão da heurística proposta no próximo capítulo.

## 4.1 Análise de distribuição dos modos de predição

Esta seção apresenta a distribuição dos modos de particionamento utilizados na codificação de mapas de profundidade do 3D-HEVC. Tal estudo compõe uma etapa preliminar da análise foco desta seção, visando apresentar justificativas para a realização da mesma. As configurações dos experimentos foram definidas de acordo com o documento das Condições Comuns de Teste (*Common Test Conditions* - CTC) do 3D-HEVC (RUSANOVSKYY; MULLER; VETRO, 2013). O software de referência do 3D-HEVC, o *3D-HEVC Test Model* (HTM), em sua versão 15.0 (CHEN et al., 2015) foi utilizado nesta análise para codificar e extraír dados de codificação de sete sequências de vídeo: *Balloons*, *Kendo* e *Newspaper* com resolução de 1024x768 pixels e *GT\_Fly*, *Poznan\_Hall2*, *Poznan\_Street* e *Undo\_Dancer*, com resolução de 1920x1088 pixels. Para cada sequência, 48 quadros - com três vistas de textura e seus respectivos mapas de profundidade - foram codificados, utilizando o menor e o maior Parâmetro de Quantização (*Quantization Parameter* - QP) definido na CTC; ou seja: QP 25 e 40 para o canal de textura e QP 34 e 45 para os mapas de profundidade. Visto que a análise apresentada neste capítulo é realizada na codificação de mapas de profundidade, quando for mencionado o QP de codificação, estará referindo-se ao QP utilizado na profundidade. Destaca-se que os dados coletados durante as simulações foram analisados com a ferramenta *3D-HEVC Analyzer*, detalhada no Apêndice A desta dissertação de mestrado.

Uma vez definida as configurações do experimento, a seguir serão discutidos os resultados da análise descrita nesta seção. A Figura 23 demonstra a ocorrência de cada modo de particionamento de CUs em PUs para as sete sequências de vídeo listadas acima. A Figura 23 (a) apresenta os resultados da análise efetuada no codificador utilizando QP 34, enquanto que a Figura 23 (b) ilustra estas informações considerando QP 45.

Como é possível observar, o particionamento 2Nx2N possui uma taxa de

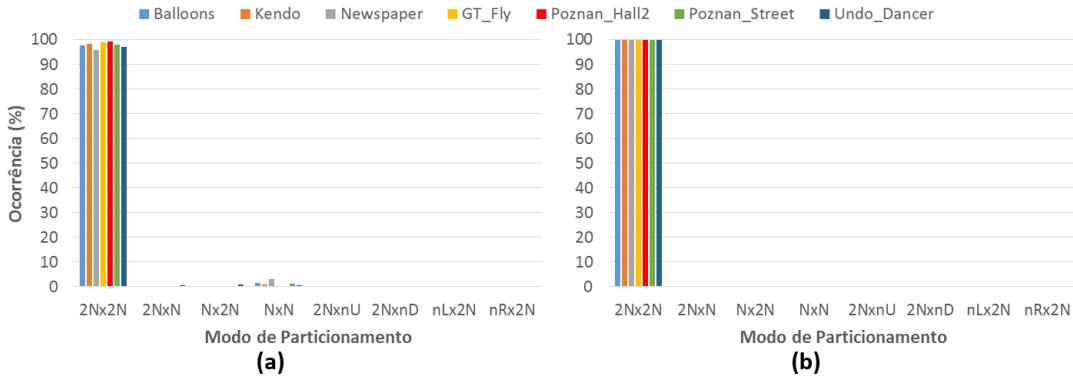


Figura 23: Distribuição dos modos de particionamento de PUs de mapas de profundidade para diversas sequências de teste utilizando (a) QP 34 e (b) QP 45

ocorrência muito superior aos demais modos, atingindo mais de 90% em todos os casos listados. Considerando apenas a análise realizada com QP 45, observa-se que a taxa do modo 2Nx2N é muito próxima a 100% em todos os vídeos avaliados.

A visualização da distribuição para os demais modos na Figura 23, em ambos cenários, é prejudicada devido à discrepância existente entre as taxas de ocorrência. Para uma melhor observação da distribuição, a Figura 24 ilustra os mesmos dados da Figura 23, entretanto apresentando-os em escala logarítmica.

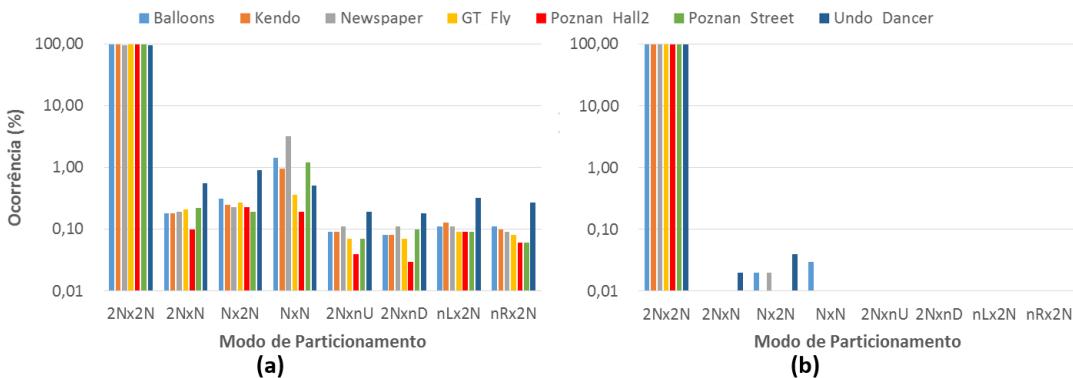


Figura 24: Distribuição em escala logarítmica dos modos de particionamento de PUs de mapas de profundidade para diversas sequências de teste utilizando (a) QP 34 e (b) QP 45

Através da Figura 24 (a), observa-se que, em quatro das sete sequências de vídeo, a ocorrência do particionamento NxN é superior a todos os demais modos apresentados no gráfico (exceto o 2Nx2N), chegando a atingir uma taxa de ocorrência bem superior dos demais no vídeo *Newspaper*. Além disso, observa-se na Figura 24 (a) que a taxa de ocorrência dos modos de particionamento assimétricos (2NxU, 2NxD, nLx2N e nRx2N) é inferior à dos demais, não ocorrendo em mais que 1% dos casos.

Por fim, observa-se na Figura 24 (b) que a taxa de ocorrência de todos os modos (exceto o 2Nx2N) é praticamente zero ao utilizar QP 45 na codificação de mapas de profundidade.

Visando facilitar a avaliação da taxa de ocorrência dos modos assimétricos (2NxnU, 2NxnD, nLx2N e nRx2N) e retangulares (2NxN, Nx2N, 2NxnU, 2NxnD, nLx2N e nRx2N), a Figura 25 ilustra os dados apresentados nos gráficos presentes nas Figuras 23 e 24, entretanto agrupando-os conforme suas respectivas categorias.

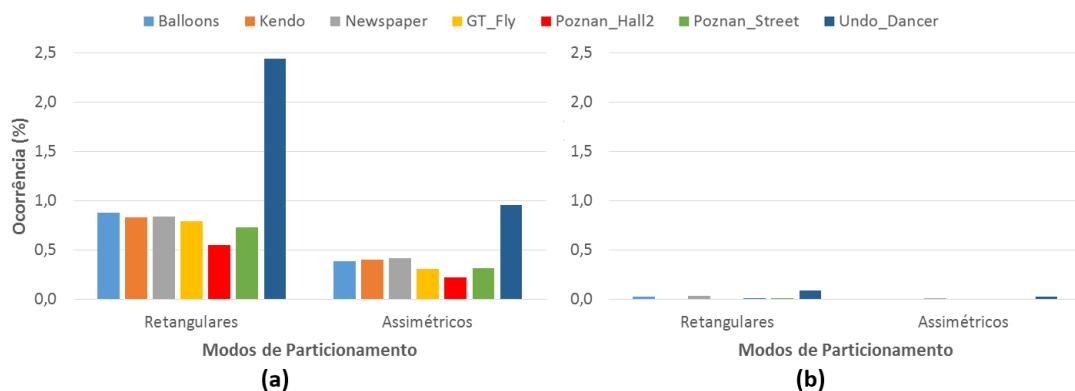


Figura 25: Distribuição dos modos de particionamento de PUs classificados em retangulares e assimétricos de mapas de profundidade para diversas sequências de teste utilizando (a) QP 34 e (b) QP 45

Observa-se na Figura 25 (a) que, quando codificados com QP 34, a incidência de modos retangulares é inferior a 1% em praticamente todos os vídeos avaliados. Além disso, nota-se que este valor é ainda menor quando considerados apenas os modos de particionamento assimétricos. Somando-se a isto, avalia-se na Figura 25 (b) que, em nenhuma sequência de vídeo codificada com QP 45, a ocorrência de modos assimétricos ou retangulares atingiu 0,5% de ocorrência.

Através dos gráficos ilustrados nas Figuras 23, 24 e 25, conclui-se que a taxa de ocorrência do modo de particionamento 2Nx2N é muito superior à dos demais. Isto se deve a diversos fatores, tal como: o uso de mais de uma PU por CU aumenta o custo de codificação do bloco, visto que mais dados laterais deverão ser transmitidos ao decodificador; as heurísticas de codificação implementadas no 3D-HEVC, tal como o DQP e o QTL, preterem a seleção - e até mesmo o teste - de PUs além da 2Nx2N e os mapas de profundidade tendem a possuir amplas áreas homogêneas, que são melhor codificadas com tamanhos de blocos maiores.

Por fim, outro fator que reduz ainda mais a seleção dos modos de particionamento retangulares é o uso exclusivo do modo 2Nx2N na implementação das ferramentas de codificação que apresentam baixíssimo custo de codificação, tal como o *Skip* e o *DIS*. Visto que estas ferramentas efetuam uma predição muito eficiente das amostras,

utilizando poucos bits para tanto, a ocorrência destes tende a ser muito alta, elevando ainda mais ocorrência do modo  $2Nx2N$ . Este fato é comprovado na Figura 26, onde são apresentadas as taxas de ocorrência dos modos *Skip*, *Merge* e *DIS*, utilizando os QPs (a) 34 e (b) 45, considerando qualquer modo de particionamento de PUs. Conforme elucidado no Capítulo 2, o modo *Skip* é um caso especial do *Merge*, entretanto, os gráficos ilustrados na Figura 26 apresentam a taxa de ocorrência do modo *Merge* quando este não é *Skip*, por isto é denominado nos gráficos ilustrados na Figura 26 como “*Merge Apenas*”.

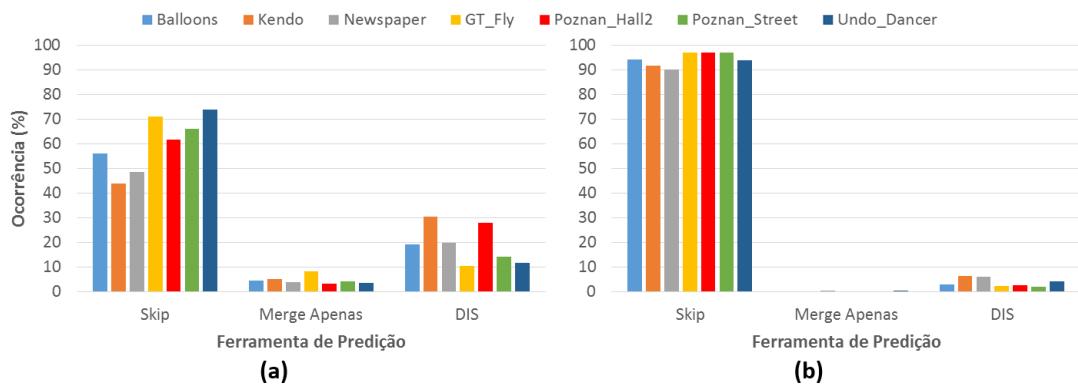


Figura 26: Taxa de ocorrência das ferramentas de predição *Skip*, *Merge Apenas* e *DIS*, utilizando QP (a) 34 e (b) 45

Percebe-se, através dos gráficos ilustrados na Figura 26, que a taxa de ocorrência do modo *Skip* é muito superior à dos demais. Considerando o QP 34, no pior caso (menor taxa de ocorrência), o *Skip* é selecionado em cerca de 40% das CUs. No melhor caso, a ocorrência deste modo atinge mais de 70%. Por fim, destaca-se que entre todas as sequências avaliadas, o modo *Skip* apresenta uma taxa de ocorrência média de 60%. É importante salientar que a taxa de ocorrência do modo *DIS* varia entre 10% a 30% considerando o QP 34. Estes fatos demonstram a relevância destes dois modos no processo de codificação de mapas de profundidade do 3D-HEVC, mesmo no cenário onde o menor valor de QP (recomendado nas CTCs) é utilizado, o que indica uma condição onde a qualidade do vídeo codificado é prioridade. Salienta-se, por fim, que a ocorrência do modo *Merge* (Apenas) é de, no máximo, 10% em todas as sequências analisadas.

Considerando os dados obtidos através dos experimentos realizados utilizando QP 45, observa-se que a taxa de ocorrência do modo *Skip* é muito superior à dos demais modos avaliados, com taxa de ocorrência superior à 90% em todas as sequências de vídeo analisadas, chegando a atingir 99% em alguns casos. Desta forma, observa-se um impacto ainda maior do modo *Skip* neste cenário. Além disso, destaca-se que em todos os casos avaliados utilizando QP 45, a ocorrência de um dos modos (*Skip*

ou DIS), é superior a 96% em todos os vídeos avaliados, demonstrando a relevância ainda maior destas duas ferramentas de codificação quando valores mais elevados de QP são considerados, indicando prioridade para a redução na taxa de bits do vídeo codificado.

É possível concluir, a partir destas análises, que a elevada taxa de ocorrência do modo 2Nx2N, tal qual apresentado na Figura 23 e na Figura 24, é função, principalmente dos modos *Skip*, DIS e *Merge*, que são aplicados apenas em PUs 2Nx2N.

As observações apresentadas nesta seção motivaram a análise RDC (*Rate-Distortion-Complexity*) apresentada na próxima seção, bem como nortearam a heurística descrita no próximo capítulo. A próxima seção apresenta a avaliação do impacto da remoção dos modos de particionamento retangulares e assimétricos da codificação de mapas de profundidade. Como estes modos são pouco usados, esta avaliação foi desenvolvida para analisar a redução de complexidade e de eficiência de compressão quando estes modos são simplesmente excluídos do processo de codificação.

## 4.2 Análise RDC com a remoção de modos de predição

Assim como na análise de distribuição dos modos apresentada na seção anterior, a análise da remoção dos modos assimétricos e retangulares foi realizada codificando as sequências de vídeo *Balloons*, *Kendo* e *Newspaper* na resolução de 1024x768 pixels e *GT\_Fly*, *Poznan\_Hall2*, *Poznan\_Street* e *Undo\_Dancer* na resolução de 1920x1088 pixels, de acordo com as configurações especificadas na CTC e utilizando os quatro QPs definidos no documento (25/34, 30/39, 35/42 e 40/45 - QP de textura/QP de profundidade) e três vistas (com seus respectivos mapas de profundidade) para cada sequência. Além disso, o total de quadros codificados para cada vídeo também foi definido conforme especificado na CTC.

Duas versões modificadas do software de referência do 3D-HEVC foram geradas, a fim de realizar esta análise. A primeira corresponde ao HTM versão 16.0 com os modos de particionamento assimétricos na codificação de mapas de profundidade desabilitados (SA-HTM). A segunda corresponde ao HTM versão 16.0 com os modos de particionamento retangulares desabilitados (SR-HTM). É importante destacar que a remoção dos modos retangulares resulta, naturalmente, na remoção dos modos assimétricos, visto que o conjunto de modos assimétricos está contido dentro do conjunto de modos retangulares. Desta forma, três implementações do software de referência foram utilizadas nesta análise: o HTM versão 16.0 original, o SA-HTM e o SR-HTM.

Todo o experimento foi realizado em um computador Intel® Core™ i7 4770 (3,4GHz), com 16GB de memória RAM, DDR3 de 1333MHz, e sistema operacional Xubuntu (baseado no Ubuntu) versão 15.04.

Os resultados apresentados para esta análise consideram a taxa de bits (*bitrate*), qualidade de imagem e complexidade do processo de codificação. Os dados de *bitrate* e qualidade de imagem são correlacionados através da métrica Bjøntegaard Delta rate (BJONTEGAARD, 2001). Além disso, neste trabalho a redução de complexidade é avaliada através da redução do tempo de codificação.

Visando facilitar a compreensão da métrica BD-rate, define-se  $bs_{ref}$  como o *bits-tream* gerado pela codificação de uma sequência de vídeo realizada pelo HTM e  $bs_{mod}$  como o *bitstream* gerado pela codificação de uma sequência de vídeo realizada pelo HTM com algum tipo de modificação (SA-HTM ou SR-HTM, por exemplo). Assim, um acréscimo percentual em BD-rate significa o percentual de aumento na taxa de bits de  $bs_{mod}$  em relação à de  $bs_{ref}$  para que ambos vídeos reconstruídos possuam a mesma qualidade objetiva de imagem. Portanto, conclui-se que um aumento em BD-rate demonstra que as modificações efetuadas no HTM geraram um impacto negativo na eficiência do codificador.

A redução de complexidade é dada observando a diminuição de tempo de codificação do processo. O cálculo percentual da redução de tempo (% $T_{red}$ ) é obtido através da Equação 1, onde  $T_{ref}$  significa o tempo de codificação obtido com o HTM e  $T_{mod}$  significa o tempo processamento obtido com o SA-HTM ou SR-HTM.

$$\%T_{red} = \left( \frac{T_{ref} - T_{mod}}{T_{ref}} \right) \times 100 \quad (1)$$

Os resultados de redução de complexidade e impacto na eficiência de codificação para o SA-HTM são apresentados na Tabela 1. A coluna *Textura 0* indica a variação percentual em BD-rate para a vista 0 do canal de textura, ao passo que as colunas *Textura 1* e *Textura 2* indicam este mesmo dado para as vistas 1 e 2, respectivamente. A coluna *Apenas Textura* indica a média da variação percentual em BD-rate das três vistas do canal de textura. Embora a remoção dos modos de particionamento tenha sido realizada exclusivamente no canal de profundidade, esta alteração interfere na eficiência de codificação das vistas dependentes de textura. Isto ocorre pois, conforme visto no Capítulo 3, determinadas ferramentas de codificação de vistas dependentes utilizam mapas de profundidade previamente codificados visando obter informações de profundidade da imagem.

Somando-se a isto, a remoção dos modos assimétricos afeta diretamente a qualidade dos mapas de profundidade decodificados, os quais não são exibidos ao espectador, mas utilizados no processo de síntese de vistas. Desta forma, o impacto das modificações é avaliado comparando as taxas de bits totais (textura + profundidade) geradas pelos codificadores HTM e SA-HTM, além de observar a distorção de imagem das vistas sintetizadas a partir da decodificação do *bitstream* gerado pelo SA-HTM, comparando-as com a perda da qualidade das vistas sintetizadas a partir da

Tabela 1: Resultados de simulação da análise RDC para o SA-HTM

Vídeo	Textura 1 (BD-rate)	Textura 2 (BD-rate)	Textura 3 (BD-rate)	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Prof.)
<i>Balloons</i>	0,000%	-0,106%	-0,002%	-0,011%	0,003%	0,1%	1,5%
<i>Kendo</i>	0,000%	-0,063%	-0,035%	-0,025%	0,022%	0,5%	1,3%
<i>Newspaper</i>	0,000%	-0,081%	0,013%	-0,008%	0,044%	0,8%	1,4%
<i>GT_Fly</i>	0,000%	0,043%	0,161%	0,021%	0,014%	0,1%	1,2%
<i>Poznan_Hall2</i>	0,000%	-0,085%	-0,048%	-0,020%	0,045%	0,5%	1,5%
<i>Poznan_Street</i>	0,000%	0,005%	0,056%	0,015%	0,043%	-0,1%	-0,3%
<i>Undo_Dancer</i>	0,000%	0,141%	0,039%	0,025%	0,115%	-0,1%	0,4%
Média 1024x768	0,000%	-0,084%	-0,008%	-0,015%	0,023%	0,5%	1,4%
Média 1920x1088	0,000%	0,026%	0,052%	0,010%	0,054%	0,1%	0,7%
Média	0,000%	-0,021%	0,026%	0,000%	0,041%	0,3%	1,2%

decodificação do *bitstream* gerado pelo HTM. A correlação entre a taxa de bits total gerada pelos codificadores HTM e SA-HTM, juntamente com a qualidade de imagem das vistas sintetizadas, fornece o BD-rate para vistas sintetizadas (*Apenas Síntese*). Este processo é apresentado detalhadamente em (RUSANOVSKYY; MULLER; VETRO, 2013). Por fim, a Tabela 1 apresenta a redução de tempo total de codificação (*Total*), além de apresentar a redução de tempo de codificação apenas dos mapas de profundidade (*Prof.*).

Como é possível perceber através dos dados apresentados na Tabela 1, a remoção dos modos assimétricos da codificação de mapas de profundidade gerou um ganho percentual máximo, em termos de BD-rate para o canal de textura, de apenas 0,025%. Além disso, considerando a média dos valores obtidos na codificação das sequências de vídeos de resolução 1024x768 *pixels*, observa-se uma redução de 0,015% em BD-rate. Em outras palavras, a remoção destes modos aprimorou a eficiência do codificador neste cenário. Por fim, avaliando-se a média entre todos as sequências analisadas, não se observa nenhum acréscimo ou redução em termos de BD-rate para o canal de textura.

Ao observar os resultados de eficiência de codificação do SA-HTM para as vistas sintetizadas na Tabela 1, percebe-se um acréscimo em BD-rate médio de apenas 0,023%, considerando vídeos com a resolução 1024x768 *pixels*. Além disso, considerando a média entre todas as sequências analisadas, a remoção dos modos assimétricos proporcionou uma perda na eficiência de codificação para as vistas sintetizadas (acréscimo de 0,041% em BD-rate).

Em relação à redução do tempo total de codificação, observa-se que, na média,

Tabela 2: Resultados de simulação da análise RDC para o SR-HTM

Vídeo	Textura 1 (BD-rate)	Textura 2 (BD-rate)	Textura 3 (BD-rate)	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Prof.)
<i>Balloons</i>	0,000%	-0,022%	0,092%	0,022%	-0,007%	2,8%	6,7%
<i>Kendo</i>	0,000%	-0,006%	-0,083%	-0,020%	-0,011%	3,0%	6,9%
<i>Newspaper</i>	0,000%	-0,088%	0,031%	-0,008%	0,108%	3,0%	6,0%
<i>GT_Fly</i>	0,000%	0,067%	0,131%	0,014%	0,040%	2,9%	7,1%
<i>Poznan_Hall2</i>	0,000%	-0,094%	-0,038%	-0,035%	0,154%	1,9%	4,6%
<i>Poznan_Street</i>	0,000%	-0,027%	0,088%	0,004%	0,068%	2,2%	4,5%
<i>Undo_Dancer</i>	0,000%	0,147%	-0,009%	0,026%	0,068%	2,5%	6,6%
<i>Média 1024x768</i>	0,000%	-0,038%	0,013%	-0,002%	0,030%	2,9%	6,6%
<i>Média 1920x1088</i>	0,000%	0,023%	0,043%	0,003%	0,124%	2,3%	5,7%
<i>Média</i>	0,000%	-0,003%	0,030%	0,001%	0,084%	2,6%	6,1%

a remoção dos modos assimétricos para os mapas de profundidade proporcionou um ganho inexpressivo, de apenas 0,3%. Somando-se a isto, observa-se um acréscimo percentual no tempo de codificação (redução negativa) em algumas sequências de vídeo. Portanto, conclui-se que, em dados cenários, esta solução conduz a resultados indesejáveis no que diz respeito à redução de complexidade.

Por fim, observa-se, através dos resultados apresentados na Tabela 1, que, em média, o SA-HTM é 1,2% menos complexo que o HTM. Assim como ocorreu nos resultados apresentados para a redução do tempo total de codificação, a sequência Poznan\_Street obteve um resultado negativo em relação à redução de complexidade (0,3%). Além disso, observa-se que, na média, a remoção dos modos assimétricos demonstrou-se duas vezes mais eficientes nas sequências de resolução 1024x768 quando comparada às sequencias 1920x1088, em termos de redução de tempo de execução na codificação dos mapas de profundidade.

A Tabela 2 apresenta os resultados da análise RDC para a remoção dos modos de particionamento retangulares da codificação de mapas de profundidade. Assim como na Tabela 1, os dados de redução/acríscimo na eficiência de codificação do canal de textura, tal como das vistas sintetizadas são apresentados. Além disso, resultados de redução de tempo total de codificação, tal como apenas do codificador de profundidade, serão exibidos na Tabela 2.

Observa-se na Tabela 2 que o SR-HTM apresentou uma perda média de 0,001% na eficiência de codificação de textura. Entretanto, os valores de BD-rate para este cenário são discrepantes, variando entre -0,035% e 0,026%. Com isto, conclui-se que a eficiência de codificação de textura pode sofrer impactos tanto negativos, quanto

Tabela 3: Comparação entre os resultados médios da análise RDC de SA-HTM e SR-HTM

Implementação	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Profundidade)
SA-HTM	0,000%	0,041%	0,3%	1,2%
SR-HTM	0,001%	0,084%	2,6%	6,1%

positivos, em termos de BD-rate com a remoção dos modos retangulares de particionamento do codificador de profundidade do 3D-HEVC. No entanto, em todos os casos, os impactos são muito pequenos e próximos de zero.

A codificação de vistas sintetizadas sofreu um acréscimo médio de 0,084% em BD-rate, chegando a 0,124% nas sequências de resolução 1920x1088. Percebe-se que, em alguns casos, a remoção dos modos assimétricos proporcionou uma redução em termos de BD-rate (sequências *Balloons* e *Kendo*).

Em relação ao tempo total de codificação, ao contrário do SA-HTM, todas as sequências de teste apresentaram resultados positivos neste quesito ao serem codificadas com o SR-HTM. Observa-se que, considerando apenas os vídeos de resolução 1024x768, a redução média do tempo total de codificação é de aproximadamente 3%.

A Tabela 2 também apresenta que o tempo de codificação dos mapas de profundidade foi reduzido em todos os casos ao desativar os modos de predição retangulares da codificação de mapas de profundidade. Em média, o SR-HTM atingiu 6,1% de redução de complexidade, chegando a 6,6% avaliando apenas vídeos de resolução 1024x768.

A Tabela 3 apresenta uma comparação entre os resultados médios de impacto na eficiência de codificação e na redução do tempo de execução obtidos com as implementações SA-HTM e SR-HTM.

Como é possível perceber, tanto o SA-HTM quanto o SR-HTM apresentaram, em média, impactos pouco significativos na eficiência do codificador de textura do 3D-HEVC, tanto na codificação de quadros de textura, quanto de vistas sintetizadas. Desta forma, conclui-se que a remoção dos modos assimétricos e retangulares não acarreta em impactos significativos na eficiência de codificação do 3D-HEVC. Por fim, verifica-se uma redução de tempo na codificação de mapas de profundidade aproximadamente cinco vezes maior no SR-HTM quando comparado ao SA-HTM. Isto ocorre devido a maior restrição no número de modos de particionamento imposta pelo SR-HTM quando comparado ao SA-HTM.

Embora as implementações SA-HTM e SR-HTM evitam o teste de 50% a 75% dos modos de particionamento estipulados pelo 3D-HEVC, observa-se que a redução de complexidade percebida por estes varia apenas entre 1,2% a 6,1%. Isto se deve às

heurísticas de redução de complexidade implementadas no HTM, as quais evitam o teste de determinados modos de particionamento - como o QTL, por exemplo - dada algumas condições. Com isto, justifica-se a baixa redução de complexidade obtida frente ao número de modos desativados.

Considerando todos os resultados apresentados, é possível concluir que os modos de particionamento assimétricos e retangulares, além de serem pouco usados, também contribuem pouco para a eficiência de codificação, ao ponto de sua remoção gerar impacto quase nulo em BD-rate e muito baixo em termos de redução no tempo de execução. Essa observação reforça a conclusão apresentada na seção anterior, qual seja: o modo de particionamento mais importante para a codificação dos mapas de profundidade é o 2Nx2N.

Destaca-se que, tanto para o SA-HTM, quanto o SR-HTM, a eficiência de codificação do canal de textura é impactada pelas modificações efetuadas no codificador de mapas de profundidade devido à forma em que as ferramentas de codificação de vistas dependentes são empregadas. Conforme descrito no Capítulo 3, algumas destas utilizam mapas de profundidade reconstruídos (codificados e decodificados) visando obter parâmetros e informações de movimento. Com isto, alterações efetuadas no codificador de mapas de profundidade acarretarão em modificações nestes quadros reconstruídos, os quais serão utilizados pelas ferramentas de codificação de vistas dependentes. Desta forma, justifica-se as variações em BD-rate apresentadas no canal de textura das vistas 1 e 2 (independentes).

Além disso, é importante salientar que alguns cenários perceberam um aumento na eficiência de compressão no codificador de textura e/ou vistas sintetizadas (redução em BD-rate). Isto é possível em função da forma como o 3D-HEVC efetua algumas das decisões de modo, considerando apenas dados locais, como distorção e estimativa da taxa de bits do bloco que está sendo codificado. Portanto, nestes casos, o codificador não verifica o resultado global de suas decisões. Logo, uma alteração na estrutura da decisão de modo do codificador pode, em alguns cenários, conduzir o codificador a melhores resultados globais em termos de eficiência de codificação.

### **4.3 Considerações Finais do Capítulo**

Este capítulo apresentou uma análise sobre o processo de codificação dos mapas de profundidade do padrão HEVC, que serviram de base para o desenvolvimento da heurística desenvolvida neste trabalho, que é apresentada no próximo capítulo. A primeira análise avaliou a distribuição dos modos de particionamento na codificação de mapas de profundidade. Esta análise demonstrou que o modo 2Nx2N é o modo de partição mais amplamente utilizado, ocorrendo em mais de 90% em todos os casos avaliados. Além disso, essa análise mostrou que, no melhor caso, os modos retan-

gulares ocorrem no máximo 2,5% das vezes, ao passo que os modos assimétricos ocorrem em apenas 1% das vezes.

Além disso, esta análise também mostrou que a elevada taxa de ocorrência dos modos *Skip*, *Merge* e *DIS* na codificação dos mapas de profundidade é definidora da dominância do uso do modo de partição  $2Nx2N$ , dados que estes modos de codificação são aplicados apenas em partições  $2Nx2N$ . Estes modos, em conjunto, são escolhidos entre 72% e 99% das vezes (91% na média), o que é quase a totalidade do uso do modo de partição  $2Nx2N$ . Finalmente, entre os três modos, o mais utilizado é o *Skip* (79% na média), seguido pelo *DIS* (11% em média) e o *Merge* (2% em média), considerando os dois QPs avaliados.

A seguir, foi realizada uma análise para avaliar o impacto que a exclusão dos modos assimétricos e retangulares poderiam trazer na relação entre taxa de bits, qualidade de imagem e complexidade da codificação de mapas de profundidade do 3D-HEVC. Esta análise é importante, uma vez que os modos menos usados podem ser responsáveis por ganhos importantes de codificação. Os resultados desta avaliação mostraram que os impactos na eficiência de codificação são inexpressivos com a exclusão destes modos de particionamento, tanto para textura (0,001% na média), quanto para as vistas sintetizadas (0,084% na média). Além disso, a redução do tempo de codificação também foi baixa, com uma média de 2,6% no codificador global e 6,1% no codificador de mapas de profundidade. Assim é possível concluir que estes modos, de fato, pouco contribuem para a codificação dos mapas de profundidade, tanto do ponto de vista de eficiência de codificação, quanto do ponto de vista de tempo de execução.

Com os resultados apresentados neste capítulo, é possível concluir que o desenvolvimento de heurísticas para a redução do tempo de execução da codificação dos mapas de profundidade deve estar focada nos modos de partição  $2Nx2N$ , em especial nos modos *Skip* e *DIS*, que são os mais usados neste cenário e que apresentam as maiores contribuições para a eficiência de codificação.

Essa conclusão foi a principal motivação para o desenvolvimento da solução apresentada no próximo capítulo: uma heurística de decisão antecipada para o 3D-HEVC, denominada *Early Skip/DIS*.

## 5 HEURÍSTICA *EARLY SKIP/DIS*

Conforme descrito anteriormente, diversos modos e ferramentas compõem o processo de codificação de vídeos do 3D-HEVC. Dentre estes, dois destacam-se pelo fato de não transmitirem informação residual: os modos *Skip* e *DIS*, sendo estes os modos mais importantes no processo de codificação de mapas de profundidade tendo em vista a alta taxa ocorrência destes modos, conforme visto no capítulo anterior.

Assim, esse trabalho propõe uma heurística que avalia primeiramente os modos *Skip* e/ou *DIS* e, considerando uma decisão local quanto a eficiência de codificação, decide se os demais modos devem ou não ser avaliados. Essa heurística é denominada *Early Skip/DIS* e é capaz de gerar importantes reduções de complexidade a um baixo custo na eficiência de codificação.

A ideia básica é definir um critério local e simplificado de avaliação do custo taxadistorção (*RD* - *Rate-Distortion*) e, a partir deste custo *RD*, definir *thresholds* para decidir antecipadamente o uso dos modos *Skip* e *DIS*. Quanto menor for o custo *RD* gerado pelos modos *Skip* ou *DIS*, maior será a probabilidade desses modos ocorrerem. Assim, para valores *RD* abaixo dos *thresholds* definidos, apenas os modos *Skip* e *DIS* são avaliados. A heurística também considera que os *thresholds* devem ser adaptativos às características do vídeo (movimentação, textura, iluminação, etc.), como forma de ampliar a sua eficiência.

No 3D-HEVC, assim como no HEVC, o custo *RD* é basicamente calculado conforme definido na Equação 2, onde *dist* significa a distorção (erro de predição) causado pelo modo. Diversas funções podem ser utilizadas para definir a distorção de um determinado modo, tal como a soma do erro quadrático (*Sum of Square Error* - *SSE*) e a soma das diferenças absolutas (*Sum of Absolute Difference* - *SAD*). Além disso, na Equação 2 a variável *bits* representa a estimativa da quantidade de informação necessária para codificar o modo em questão. Por fim, a variável lambda ( $\lambda$ ) na equação significa um multiplicador de Lagrange, o qual varia de acordo com determinadas configurações de codificação (QP, por exemplo), visando priorizar a qualidade de imagem ou compressão do bloco. O custo *RD* dos modos utilizados na heurística desenvolvida é obtido diretamente no HTM (ROSEWARNE et al., 2015).

$$custoRD_{modo} = dist + (bits \times \lambda) \quad (2)$$

Esse capítulo irá apresentar, em detalhes, a heurística desenvolvida. Inicialmente, será apresentada uma análise inicial de ocorrência dos modos *Skip* e *DIS* em função do custo RD gerados por estes modos. Em seguida, será descrita a heurística *Early Skip/DIS* para o codificador de mapas de profundidade do 3D-HEVC. É importante destacar que, tendo em vista os resultados inexpressivos alcançados pela remoção dos modos de particionamento assimétricos e/ou retangulares do processo de codificação de mapas de profundidade, estes modos não foram desabilitados nos experimentos realizados neste capítulo.

## 5.1 Análise de ocorrência dos modos *Skip* e *DIS* em função do custo RD

Este experimento foi realizado utilizando o software de referência do 3D-HEVC, o HTM versão 16.0 (CHEN et al., 2015), e as configurações dos experimentos foram definidas de acordo com a CTC do 3D-HEVC (RUSANOVSKY; MULLER; VETRO, 2013). Sete sequências de vídeo foram utilizadas: *Balloons*, *Kendo* e *Newspaper*, com 1024x768 pixels e *GT\_Fly*, *Poznan\_Hall2*, *Poznan\_Street* e *Undo\_Dancer* com 1920x1088 pixels. Para cada sequência, oito quadros - com três vistas de textura e seus respectivos mapas de profundidade - foram codificados utilizando QP 30 para textura e 39 para mapas de profundidade. Utilizou-se apenas oito quadros de cada sequência nestes experimentos visto que a heurística *Early Skip/DIS*, conforme será apresentado posteriormente, avalia os oito primeiros quadros do vídeo para decidir quais valores de *thresholds* utilizar no processo de tomada de decisão antecipada.

A análise foi realizada observando a taxa de ocorrência (*TO*) dos modos *Skip* e *DIS* na codificação dos mapas de profundidade. Esta taxa é definida na Equação 3, onde  $S_M(i, d)$  é uma função que retorna o número de vezes em que o modo  $M$  (onde  $M \in \{Skip, DIS\}$ ) resultou em um custo RD  $i$ , e foi definido para a CU (de tamanho  $d$ ) que está sendo codificada. A função  $N_M(i, d)$  é similar à função  $S_M(i, d)$ , entretanto, retorna o número de vezes em que o modo  $M$  não foi definido para a CU em questão.

$$TO_M(r, d) = \frac{\sum_{i=0}^r S_M(i, d)}{\sum_{i=0}^r S_M(i, d) + \sum_{i=0}^r N_M(i, d)} \quad (3)$$

Como é possível perceber na Equação 3, para cada custo RD  $r$  e tamanho de CU  $d$ , o resultado é gerado observando-se todos os valores de zero até  $r$ . Assim, estima-se a probabilidade ocorrência dos modos *Skip* e *DIS* quando estes resultam em um custo RD  $r$ . É importante salientar que, nesta análise, os valores de  $S_M$  e  $N_M$  de cada custo RD de todos os vídeos foram somados a fim de obter uma visão geral da taxa

de ocorrência dos modos *Skip* e *DIS* dentre todas as sequências de teste avaliadas.

A Figura 27 apresenta o gráfico da taxa de ocorrência, conforme a Equação 3, dos modos (a) *Skip* e (b) *DIS* em função do custo RD gerado (considerando valores de custo RD entre zero e 10.000). Para cada gráfico, quatro curvas são apresentadas, representando os quatro tamanhos de PU possíveis para o modo de partição 2Nx2N: 64x64, 32x32, 16x16 e 8x8.

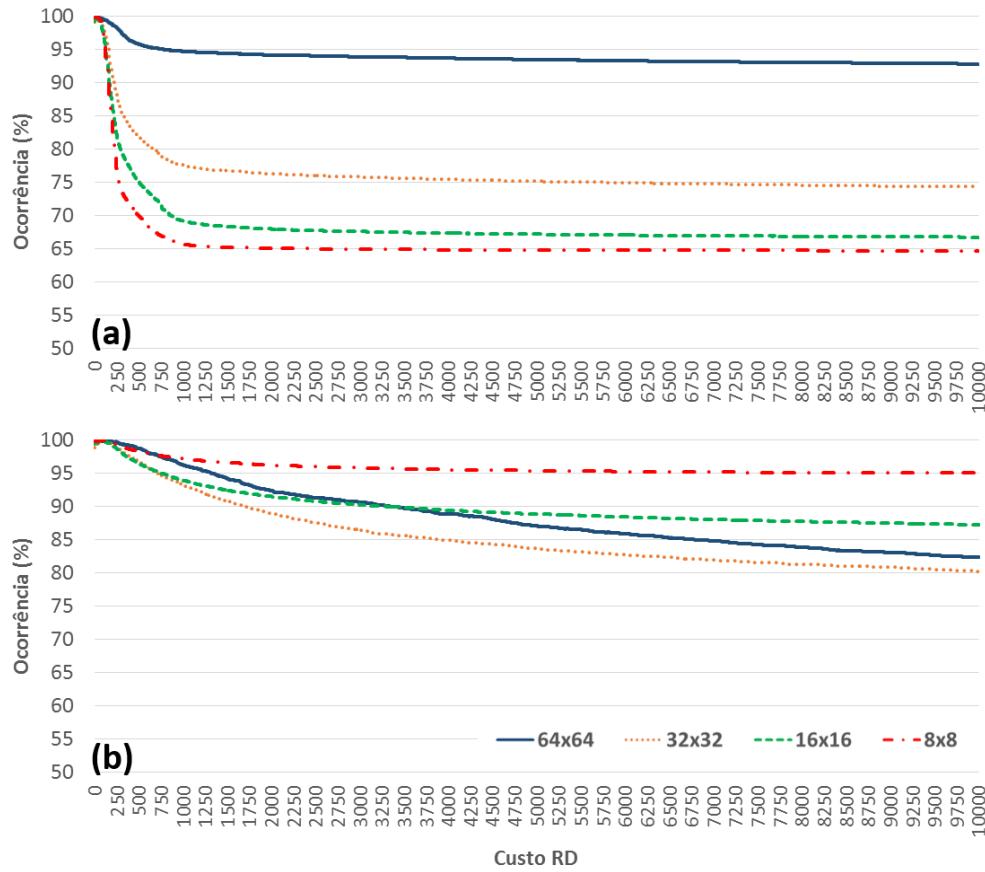


Figura 27: Taxa de ocorrência dos modos (a) *Skip* e (b) *DIS* em função do custo RD gerado

Nota-se, na Figura 27, que a taxa de ocorrência de ambos os modos é inversamente proporcional ao custo RD gerado em todos os cenários observados. Em outras palavras, há uma grande probabilidade do melhor modo ser *Skip* ou *DIS* quando um destes resulta em um custo RD baixo. Por exemplo, considerando os dados apresentados na Figura GRAFTOTAL (a), nota-se que o modo *Skip* apresenta uma taxa de ocorrência de aproximadamente 95% de ocorrência quando este resulta em um custo RD entre 750 e 1000 em blocos de tamanho 64x64. Considerando blocos de tamanho 8x8, o modo *DIS* apresenta uma taxa de ocorrência superior a 95% quando este resulta em um custo RD até 10.000.

Assim, esse trabalho teve por hipótese inicial o desenvolvimento de uma heurística

que avalie primeiramente os modos *Skip* e/ou *DIS* e, quando seus custos RD forem baixos, defina que um destes modos será utilizado, evitando a avaliação dos demais modos de codificação. Um desafio associado ao desenvolvimento desta heurística é identificar *thresholds* de RD que definirão se os demais modos deverão, ou não, serem avaliados durante o processo de codificação.

Ao observar as taxas de ocorrência apresentadas na Figura 27, é possível estimar *thresholds* fixos para esta abordagem, que, em tese, garantiriam uma determinada taxa de acerto da heurística. Por exemplo, caso defina-se um *threshold* 10.000 para a aplicação do modo *Skip* em blocos de tamanho 64x64, estima-se que a heurística acertaria mais de 90% das vezes. Entretanto, ressalta-se que as sequências de vídeo testadas para a geração do gráfico ilustrado na Figura 27 apresentam características muito distintas entre si. Desta forma, as curvas de ocorrência em função do custo RD tendem a variar muito entre uma sequência e outra.

Este fato é demonstrado na Figura 28, onde são apresentadas a taxa de ocorrência dos modos (a) *Skip* e (b) *DIS* para o vídeo *Kendo*, assim como para os modos (c) *Skip* e (d) *DIS* para o vídeo *Undo\_Dancer*. Os dados apresentados consideram os valores de custo RD de zero até 5.000.

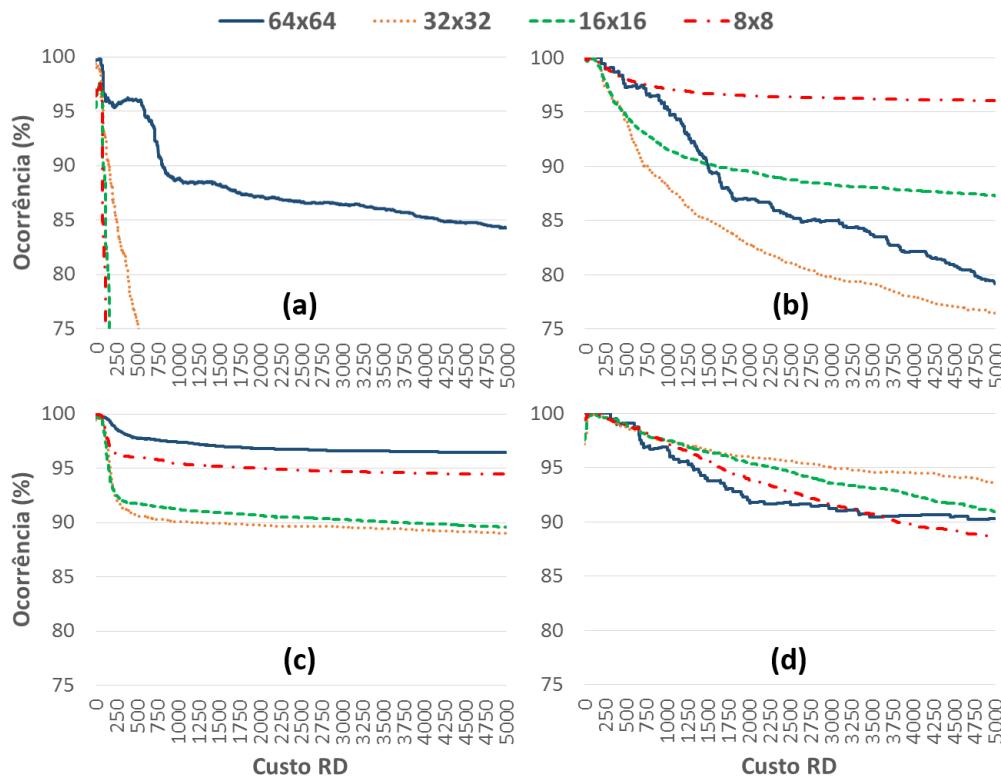


Figura 28: Análise da taxa de ocorrência dos modos (a) *Skip* e (b) *DIS* da sequência de vídeo *Kendo* e dos modos (c) *Skip* e (d) *DIS* da sequência *Undo\_Dancer*

Observa-se, na Figura 28, que a ocorrência do modo *Skip* para os blocos de tama-

nho 64x64 da sequência de vídeo *Kendo* torna-se inferior a 95% para custos RD acima de 750 aproximadamente. Entretanto, ao avaliar este mesmo dado para a sequência *Undo\_Dancer*, observa-se que até o custo RD 5.000 a taxa de ocorrência não atingiu um valor inferior a 95%. Além disso, ao considerar os demais tamanhos de bloco, a taxa de ocorrência do modo *Skip* é inferior a 75% em todos os casos na sequência de vídeo *Kendo*, a partir do custo RD 500. Não obstante, na sequência *Undo\_Dancer*, até o custo RD 5.000 a taxa de ocorrência para todos os tamanhos de bloco é sempre superior a 85%.

Ao comparar a ocorrência do modo DIS entre as duas sequências de vídeo utilizadas no experimento, observa-se que as curvas do gráfico para todos os tamanhos de bloco são completamente distintas. Por exemplo, na sequência *Kendo* a ocorrência do modo DIS para o tamanho de bloco 8x8 não atinge um valor inferior a 95% até o custo RD 5.000; entretanto, na sequência *Undo\_Dancer*, a taxa de ocorrência deste modo torna-se inferior a 95% para custos RD superiores a 1.750. Ao observar as curvas dos demais tamanhos de bloco, percebe-se que, para a sequência de vídeo *Kendo*, a taxa de ocorrência do modo DIS é inferior a 90% para o custo RD 5.000 em todos os casos. Por outro lado, o oposto é observado na sequência *Undo\_Dancer*, onde a taxa de ocorrência do modo DIS para os tamanhos de bloco 64x64, 32x32 e 16x16 é superior a 90% para o custo RD 5.000.

Conclui-se com esta análise, que a taxa de ocorrência dos modos varia de acordo com o alguns fatores, tais como o tamanho de bloco e a sequência de vídeo que está sendo codificada. Com isto, observa-se que a heurística que propõe testar apenas um, ou ambos os modos *Skip* e DIS, quando estes resultam em um custo RD baixo deve possuir *thresholds* adaptativos ao invés de fixos, visando obter melhores resultados de redução de complexidade, apresentando baixo impacto na eficiência de codificação em todos os cenários. Esta heurística é descrita detalhadamente na seção a seguir.

## 5.2 Detalhamento da heurística *Early Skip/DIS*

Considerando as análises apresentadas na seção anterior, observa-se que uma heurística *Early Skip/DIS* baseada em *thresholds* adaptativos para o codificador de mapas de profundidade do 3D-HEVC tem potencial de gerar bons resultados. Tendo isto em vista, esta seção apresenta a heurística desenvolvida para a decisão antecipada em duas partes: a primeira detalha o processo de obtenção dos *thresholds*, ao passo que a segunda parte apresenta a aplicação da heurística *Early Skip/DIS* utilizando os *thresholds* dinamicamente definidos.

### 5.2.1 Obtenção dos *Thresholds*

Uma vez que a heurística *Early Skip/DIS* utiliza dois modos (*Skip* e *DIS*) para o esquema de decisão antecipada, dois *thresholds* são necessários neste processo. O primeiro é responsável por definir o maior custo RD do modo *Skip*, o qual evitará que o codificador avalie os demais modos/ferramentas, inclusive o *DIS*. O segundo valor de *thresholds* define o maior custo RD de *Skip* ou *DIS* que evita que o codificador teste os demais modos de codificação.

A heurística *Early Skip/DIS* inicia definindo estes valores de *thresholds*, que são obtidos durante a codificação das oito primeiras Unidades de Acesso (*Access Unit* - AU) da sequência. As demais AUs são codificadas através da aplicação do algoritmo núcleo da heurística *Early Skip/DIS*, utilizando os *thresholds* encontrados. Os valores de *thresholds* são definidos, obrigatoriamente, no início da aplicação da heurística e estes valores podem ser atualizados em períodos pré-determinados de tempo a fim de garantir uma maior eficiência no processo de adaptação da heurística ao vídeo processado.

É importante destacar que dois *thresholds* são gerados para cada tamanho de PU possível, uma vez que a taxa de ocorrência dos modos *Skip* e *DIS* varia de acordo com este parâmetro. Desta forma, ao todo, utilizam-se oito *thresholds* na heurística *Early Skip/DIS*, entretanto, visto que todo o processo é o mesmo para qualquer tamanho de PU, a descrição será realizada em uma forma independente de tamanho.

O processo de obtenção dos *thresholds* é composto pela avaliação da taxa de ocorrência dos modos *Skip* e *DIS*, visando estimar a probabilidade de ocorrência destes para um dado custo RD. Desta forma, uma vez encontrado um custo RD  $r$ , o qual resulta em uma taxa de ocorrência  $X\%$  (onde  $X$  é definido pelo usuário ou aplicação), o valor do *threshold* pode ser definido como  $r$ . Assim, a heurística *Early Skip/DIS* tende a acertar aproximadamente  $X\%$  dos casos.

Duas métricas de avaliação de taxa de ocorrência são utilizadas no processo de obtenção dos *thresholds* para a heurística apresentada neste capítulo. A primeira é a métrica *TO*, definida na Equação 3 na seção 5.1. A segunda métrica utilizada no esquema proposto é denominada Taxa de Ocorrência Combinada (*TOC*), que, diferentemente de *TO*, leva em conta a taxa de ocorrência de dois modos *M1* e *M2* (onde *M1* e *M2* representam os modos *Skip* e *DIS*, respectivamente).

Na heurística proposta neste trabalho, esta métrica é utilizada após a definição do primeiro *threshold*. Assim, a heurística avalia a taxa de ocorrência do modo *M1* de zero até o primeiro *threshold* encontrado ( $r$ ), além de avaliar a taxa de ocorrência combinada dos modos *M1* e *M2* de  $r + 1$  até um segundo valor de custo RD  $s$ . De um modo geral, é possível resumir o processo dizendo que a métrica *TOC* considera a *TO* do modo *M1* de zero até um custo RD  $r$ . Entretanto, do custo RD  $r + 1$  até  $s$ , a métrica *TOC* considera a ocorrência de ambos modos *M1* e *M2*. A definição

matemática desta métrica é apresentada na Equação 4.

$$TOC_{M1,M2}(r, s, d) = \frac{\sum_{i=0}^s S_{M1}(i, d) + \sum_{i=r+1}^s S_{M2}(i, d)}{\left[ \begin{array}{l} \sum_{i=0}^s S_{M1}(i, d) + \sum_{i=r+1}^s S_{M2}(i, d) + \\ \sum_{i=0}^r N_{M1}(i, d) + \sum_{i=r+1}^s N_{M1,M2}(i, d) \end{array} \right]} \quad (4)$$

É importante destacar que esta métrica utiliza as funções previamente definidas  $S_M(i, d)$  e  $N_M(i, d)$ . Além disso, a função  $N_{M1,M2}(i, d)$  também é utilizada, a qual define o número de vezes em que o modo  $M1$  ou  $M2$  não foram definidos para uma dada PU de tamanho  $d$ . Durante o processo de definição dos *thresholds*,  $N_{M1,M2}(i, d)$  é incrementado observando o menor custo RD  $i$  obtido entre os modos  $M1$  e  $M2$ .

Os *thresholds* são definidos considerando a taxa de ocorrência dos modos *Skip* e *DIS*, visando estimar uma determinada taxa de acerto da heurística *Early Skip/DIS*. Neste cenário, taxa de acerto significa a taxa em que a heurística corretamente evitará a avaliação dos demais modos. Por exemplo, se o codificador avaliar apenas o modo *Skip*, e todos os demais modos - se avaliados - resultassem em um custo RD superior ao do *Skip*, isto é considerado um acerto neste cenário. Assim, os *thresholds* da heurística *Early Skip/DIS* são definidos considerando uma taxa de acerto alvo (*Target Hit Rate* - THR) com base na *TO* do modo *Skip* e na TOC dos modos *Skip* e *DIS*. O THR pode ser definido pela aplicação em que está inserido o codificador de acordo com requisitos de qualidade de imagem e desempenho. É importante destacar que, maiores valores de THR acarretarão em menores *thresholds* obtidos, conduzindo a menores degradações tanto na eficiência de codificação quanto na redução de complexidade. Entretanto, valores menores de THR resultarão em valores de *threshold* maiores, admitindo uma taxa de acerto menor à heurística *Early Skip/DIS*, assim, haverá maiores impactos na eficiência de codificação, porém, também haverá maior potencial para a redução de complexidade.

A fim de expandir o detalhamento sobre o THR, a Figura 29 apresenta dois gráficos de taxa de ocorrência em função do custo RD gerado pelo modo *Skip* (Figura 28 (c)), destacando quais *thresholds* seriam utilizados pela heurística *Early Skip/DIS* caso o THR selecionado fosse 95% (a) ou 90% (b). Nota-se que, caso o THR selecionado fosse 95%, os valores de *threshold* para o modo *Skip* estariam entre 0 e 250 para PUs de tamanho 16x16 e 32x32. Somando-se a isto, para blocos de tamanho 8x8 o *threshold* da heurística se encontraria entre 2.000 e 2.500. Por fim, para blocos de tamanho 64x64, o *threshold* seria maior que 5.000. Considerando o THR 90%, percebe-se que os valores de *thresholds* são superiores quando comparado ao THR 95%. O *threshold* para PUs de tamanho 32x32 estaria entre 1.250 e 1.500 ao passo que, para blocos de tamanho 16x16, o *threshold* da heurística seria aproximadamente 3.750. Uma vez descrita a constante THR, o processo de definição dos *thresholds* será descrito a seguir.

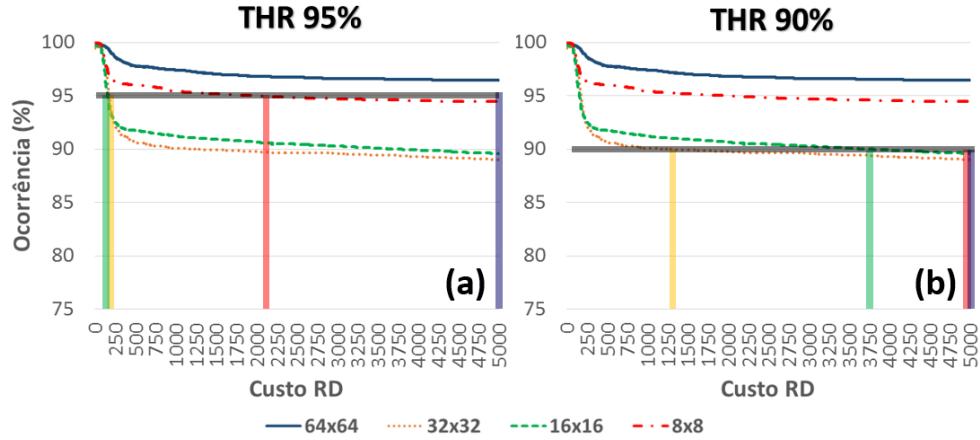


Figura 29: Exemplo de definição de *thresholds* para heurística *Early Skip/DIS* utilizando (a) THR 95% e (b) THR 90%

A Figura 30 apresenta o fluxograma de obtenção dos *thresholds* da heurística *Early Skip/DIS*, utilizando as métricas previamente definidas ( $TO$  e  $TOC$ ), além da constante  $THR$ . O primeiro *threshold* ( $TH1$ ) define o maior custo RD de *Skip* onde este modo ocorre com uma taxa  $THR$ . Além disso, o segundo *threshold* obtido ( $TH2$ ) define o maior custo RD dos modos *Skip* ou *DIS* onde um destes ocorre com uma taxa  $THR$ .

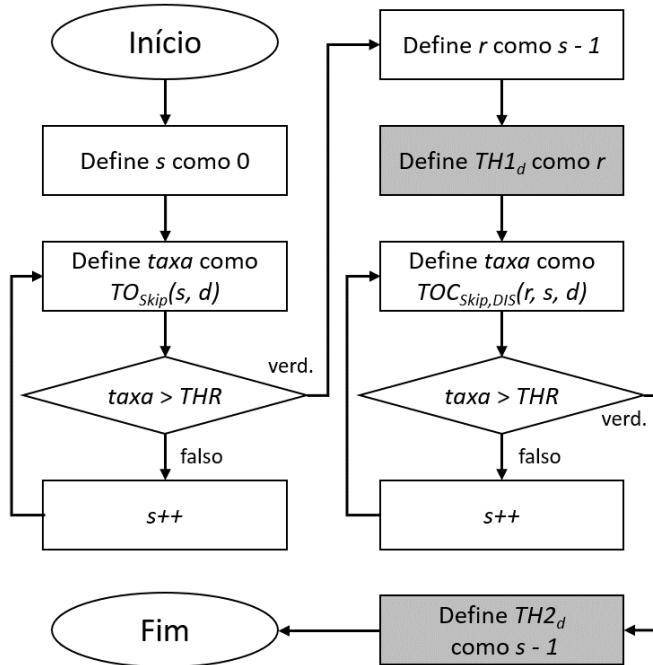


Figura 30: Fluxograma de obtenção dos *Thresholds*

Uma vez definidos os *thresholds* da heurística *Early Skip/DIS*, o codificador habilita o algoritmo do núcleo de decisão antecipada, descrito na próxima seção.

### 5.2.2 Algoritmo do núcleo da heurística *early Skip/DIS*

Assim como na seção anterior, a descrição aqui apresentada é realizada de forma independente do tamanho da PU. A Figura 31 ilustra o fluxograma do funcionamento do algoritmo do núcleo da heurística *Early Skip/DIS*, após definidos os *thresholds* TH1 e TH2. Primeiramente, avalia-se o custo RD obtido através do modo *Skip*, visando decidir se o codificador testará ou não o modo *DIS*.

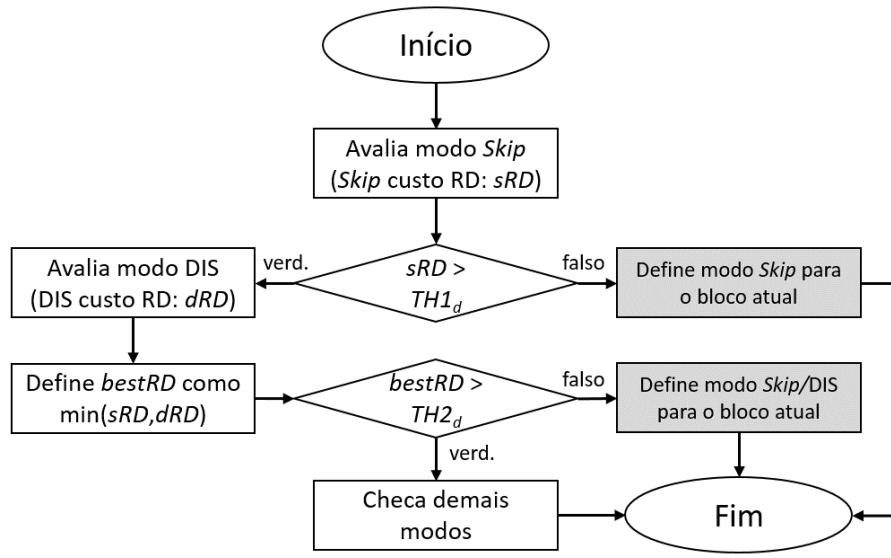


Figura 31: Fluxograma da heurística *early Skip/DIS*

Caso o custo RD do modo *Skip* ( $sRD$ ) seja menor que TH1, o codificador não avalia os demais modos, caso contrário, o modo *DIS* é avaliado e o melhor custo RD ( $bestRD$ ) é comparado com o segundo *threshold* TH2. Caso  $bestRD$  seja inferior à TH2, o modo *Skip* ou *DIS* é selecionado para o bloco atual, de acordo com qual modo obtiver o menor custo RD. Caso o valor de  $bestRD$  seja superior à TH2, o codificador avaliará os demais modos.

Por avaliar primeiramente o modo *Skip*, a heurística *Early Skip/DIS* apresentada neste capítulo prioriza a seleção deste modo em relação ao *DIS*. Isto é função da taxa de ocorrência do modo *Skip*, que é muito superior ao modo *DIS* em todos os cenários apresentados na capítulo anterior, além dos ganhos de codificação trazidos pelo modo *Skip*, que são maiores que os obtidos através do modo *DIS* (SZE; BUDAGAVI; SULLIVAN, 2014) (LEE; PARK; KIM, 2015).

## 5.3 Resultados da heurística Early Skip/DIS

A heurística *Early Skip/DIS* proposta neste trabalho foi avaliada utilizando o HTM versão 16.0, de acordo com o documento das CTC. As sete sequências de vídeo do experimento anterior foram utilizadas, considerando todos os quatro QPs estipulados

Tabela 4: Número total de quadros, taxa de exibição e tempo total das sequências de teste

Vídeo	Total de Quadros Codificados	Taxa de Quadros	Tempo Total (segundos)
Balloons	300	30	10
Kendo	300	30	10
Newspaper	300	30	10
GT_Fly	250	25	10
Poznan_Hall2	200	25	8
Poznan_Street	250	25	10
Undo_Dancer	250	25	10

nas CTC (25/34, 30/39, 35/42 e 40/45 - QP de textura/QP de profundidade). Além disso, a codificação foi realizada utilizando três vistas (com seus respectivos mapas de profundidade) para cada sequência de vídeo. Por fim, o total de quadros codificados para cada vídeo também foi definido conforme especificado na CTC. A Tabela 4 apresenta o número total de quadros codificados em cada sequência de vídeo, juntamente com a taxa de quadros (quadros por segundo) e o tempo de total de cada sequência avaliada. É importante destacar que, exceto em *Poznan\_Hall2*, todas as sequências de vídeo possuem 10 segundos de duração.

Assim como na análise apresentada no capítulo anterior, todo o experimento foi realizado em um computador Intel® Core™ i7 4770 (3,4GHz), com 16GB de memória RAM, DDR3 de 1333MHz, e sistema operacional Xubuntu (baseado no Ubuntu) versão 15.04.

Quatro valores de THR foram considerados no experimento: 99%, 95%, 90% e 80% avaliando-se a redução de complexidade em termos de redução de tempo de codificação e o impacto na eficiência de codificação em termos de BD-rate trazidos pela heurística *Early Skip/DIS*. Desta forma, cinco implementações do HTM foram utilizadas: o HTM original e quatro modificados, adaptados para executar a heurística *Early Skip/DIS* utilizando os valores de THR mencionados.

Os resultados da avaliação da heurística utilizando THR 99% são exibidos na Tabela 5. Assim como ocorreu na análise RDC para a remoção dos modos de particionamento assimétricos e retangulares, a eficiência de codificação da vista de base da textura não é impactada pelas modificações realizadas no HTM.

Observa-se que a heurística utilizando THR 99% obteve, em média, 0,012% de acréscimo em BD-rate para o canal de textura. No pior caso (sequência *GT\_Fly*), a heurística *Early Skip/DIS* gerou um impacto de apenas 0,032% em BD-rate para o canal de textura. Assim, infere-se que esta heurística utilizando THR 99% apresenta impactos pouco significativos na eficiência de codificação de quadros de textura.

A eficiência de codificação das vistas sintetizadas apresenta, em média, um

Tabela 5: Resultados de simulação da heurística *early Skip/DIS* utilizando THR 99%

Vídeo	Textura 1 (BD-rate)	Textura 2 (BD-rate)	Textura 3 (BD-rate)	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Prof.)
<i>Balloons</i>	0,000%	-0,067%	0,023%	-0,002%	0,004%	5,9%	15,7%
<i>Kendo</i>	0,000%	0,062%	-0,064%	-0,015%	-0,022%	5,4%	14,0%
<i>Newspaper</i>	0,000%	0,047%	0,001%	0,015%	0,042%	4,4%	10,2%
<i>GT_Fly</i>	0,000%	0,139%	0,250%	0,032%	0,069%	13,5%	33,6%
<i>Poznan_Hall2</i>	0,000%	0,038%	-0,014%	0,001%	0,299%	13,8%	34,0%
<i>Poznan_Street</i>	0,000%	0,125%	0,030%	0,020%	0,045%	10,9%	28,1%
<i>Undo_Dancer</i>	0,000%	0,106%	0,089%	0,031%	-0,004%	13,1%	35,0%
<i>Média 1024x768</i>	0,000%	0,014%	-0,013%	0,000%	0,008%	5,2%	13,3%
<i>Média 1920x1088</i>	0,000%	0,102%	0,089%	0,021%	0,102%	12,8%	32,7%
<i>Média</i>	0,000%	0,064%	0,045%	0,012%	0,062%	9,6%	24,4%

acréscimo de 0,062% no BD-rate. No pior caso, o acréscimo em BD-rate percebido neste cenário é de apenas 0,299% (sequência *Poznan\_Hall2*), demonstrando que a heurística desenvolvida também não gera impactos significativos na eficiência de codificação das vistas sintetizadas do 3D-HEVC, considerando THR 99%.

Os resultados de redução de tempo do codificador de profundidade apontam um ganho médio neste quesito de 24,4%. Em outras palavras, observa-se que a heurística *Early Skip/DIS* utilizando THR 99% reduziu o tempo de codificação dos mapas de profundidade no 3D-HEVC a aproximadamente um quarto do valor original. Ao observar apenas as sequências de resolução 1920x1088, percebe-se um ganho ainda maior, chegando a uma redução à aproximadamente um terço do tempo de codificação de profundidade.

Por fim, ao avaliar os resultados de redução de complexidade (tempo) total do codificador, observa-se uma redução média de 9,6%. Observando-se apenas as sequências de resolução 1920x1088, percebe-se uma maior redução de complexidade, atingindo em média 12,8% (mais de duas vezes a redução percebida nos vídeos de resolução 1024x768).

Conclui-se que a heurística *Early Skip/DIS*, utilizando THR 99%, proporciona resultados de redução de tempo codificação relevantes, impactando de forma insignificante na eficiência de codificação do 3D-HEVC.

A Tabela 6 apresenta os resultados de simulação da heurística *Early Skip/DIS* utilizando THR 95%. Assim como na Tabela 5, os dados de eficiência de codificação do canal de textura, tal como das vistas sintetizadas, são apresentados. Além disso, resultados de redução de tempo total de codificação, tal como redução de tempo do

Tabela 6: Resultados de simulação da heurística *early Skip/DIS* utilizando THR 95%

Vídeo	Textura 1 (BD-rate)	Textura 2 (BD-rate)	Textura 3 (BD-rate)	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Prof.)
<i>Balloons</i>	0,000%	-0,113%	0,073%	0,002%	-0,009%	11,1%	28,7%
<i>Kendo</i>	0,000%	0,137%	-0,091%	0,004%	0,040%	8,7%	21,8%
<i>Newspaper</i>	0,000%	-0,019%	0,053%	0,011%	0,124%	12,4%	27,9%
<i>GT_Fly</i>	0,000%	0,463%	0,426%	0,088%	0,236%	16,3%	40,3%
<i>Poznan_Hall2</i>	0,000%	-0,021%	0,084%	0,023%	1,734%	16,0%	38,8%
<i>Poznan_Street</i>	0,000%	0,012%	0,169%	0,017%	0,174%	16,1%	36,2%
<i>Undo_Dancer</i>	0,000%	0,796%	0,607%	0,182%	0,561%	16,2%	42,2%
Média 1024x768	0,000%	0,002%	0,012%	0,005%	0,052%	10,7%	25,2%
Média 1920x1088	0,000%	0,313%	0,322%	0,078%	0,676%	16,2%	39,4%
Média	0,000%	0,179%	0,189%	0,047%	0,409%	13,8%	33,7%

codificador de profundidade também estão exibidos na Tabela 5.

Os resultados de eficiência de codificação apresentados na Tabela 6 demonstram que há um acréscimo médio de 0,047% em BD-rate no canal de textura. No pior cenário avaliado (sequência *Undo\_Dancer*) observa-se um impacto negativo médio de apenas 0,182% no BD-rate. Com isto, infere-se que a eficiência da compressão de quadros de textura é um pouco mais impactada com THR 95% em comparação ao THR 99%, como esperado. Entretanto, em nenhum dos casos observados há um acréscimo em BD-rate superior a 0,8%.

Percebe-se na Tabela 6 que a eficiência de codificação das vistas sintetizadas teve um acréscimo médio de BD-rate em praticamente todas as sequências avaliadas (exceto *Balloons*). Apenas a sequência *Poznan\_Hall2* apresentou um acréscimo de BD-rate superior à 1%. Observa-se que o acréscimo médio em BD-rate para as vistas sintetizadas, considerando todas as sequências analisadas, é de 0,409%. Com isto, conclui-se que a heurística *Early Skip/DIS*, utilizando THR 80%, ainda assim apresenta um baixo impacto à eficiência de codificação das vistas sintetizadas do 3D-HEVC.

Avaliando-se apenas a redução de tempo de codificação de mapas de profundidade, observam-se ganhos muito elevados. Considerando a média de todas as sequências avaliadas, percebe-se uma redução de complexidade de 33,7%, ou seja, aproximadamente um terço do tempo de codificação é exigido. Considerando apenas a média das sequências de resolução 1920x1088, a redução de complexidade do processo atinge quase 40%, demonstrando a eficiência da heurística *Early Skip/DIS* utilizando THR 95%.

Finalmente, considerando a redução total do tempo de codificação, percebe-se que

Tabela 7: Resultados de simulação da heurística *Early Skip/DIS* utilizando THR 90%

Vídeo	Textura 1 (BD-rate)	Textura 2 (BD-rate)	Textura 3 (BD-rate)	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Prof.)
<i>Balloons</i>	0,000%	-0,042%	0,058%	0,001%	0,400%	12,8%	33,1%
<i>Kendo</i>	0,000%	0,123%	0,020%	0,013%	0,140%	11,7%	28,6%
<i>Newspaper</i>	0,000%	-0,022%	0,000%	-0,007%	0,536%	15,1%	33,6%
<i>GT_Fly</i>	0,000%	0,544%	0,558%	0,112%	0,613%	18,0%	44,6%
<i>Poznan_Hall2</i>	0,000%	0,041%	0,148%	0,047%	4,227%	16,6%	40,5%
<i>Poznan_Street</i>	0,000%	0,152%	0,236%	0,040%	0,234%	17,9%	39,7%
<i>Undo_Dancer</i>	0,000%	1,702%	1,700%	0,443%	2,395%	17,2%	44,9%
<i>Média 1024x768</i>	0,000%	0,020%	0,026%	0,002%	0,359%	13,2%	31,8%
<i>Média 1920x1088</i>	0,000%	0,610%	0,661%	0,160%	1,867%	17,4%	42,4%
<i>Média</i>	0,000%	0,357%	0,389%	0,093%	1,221%	15,6%	37,8%

o codificador apresentou ganhos de aproximadamente 13,8%, em média. Observando os dados médios, tanto as sequências de resolução 1024x768 quanto 1920x1088 apresentaram redução de complexidade superior a 10%, atingindo 16,2% nesta última. Apenas a sequência *Kendo* (resolução 1024x768 pixels) não obteve redução superior a 10%. Assim, percebe-se que a heurística *Early Skip/DIS* apresenta resultados satisfatórios em termos de redução de complexidade no HTM.

A Tabela 7 apresenta os resultados de simulação obtidos com a solução *Early Skip/DIS* utilizando a taxa de acerto alvo de 90%.

Considerando os resultados de eficiência de codificação de quadros de textura, observa-se que a heurística *Early Skip/DIS* utilizando THR 90% proporciona, em média, um acréscimo de 0,093% na eficiência de codificação de quadros de textura. Considerando apenas as sequências de resolução 1024x768, nota-se que o acréscimo médio em BD-rate é de apenas 0,002%.

Um aumento de 1,211% em BD-rate é percebido na codificação de vistas sintetizadas ao utilizar-se a heurística *Early Skip/DIS* com THR 90%. É importante destacar que este alto valor médio é atingido devido à sequência *Poznan\_Hall2*, a qual apresentou um acréscimo em BD-rate muito superior comparado com os demais experimentos. Ao desconsiderar esta sequência no cálculo da média, observa-se que o aumento em BD-rate proporcionado pelo *Early Skip/DIS* (THR 90%) é de apenas 0,72%.

Ao observar-se os dados de tempo de codificação de mapas de profundidade, verifica-se que, na média entre todas as sequências, o *Early Skip/DIS* proporciona uma redução de 37,8% neste cenário. Considerando apenas sequências de resolução

Tabela 8: Resultados de simulação da heurística *Early Skip/DIS* utilizando THR 80%

Vídeo	Textura 1 (BD-rate)	Textura 2 (BD-rate)	Textura 3 (BD-rate)	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Prof.)
<i>Balloons</i>	0,000%	0,018%	-0,018%	0,006%	0,703%	15,1%	38,5%
<i>Kendo</i>	0,000%	0,036%	0,055%	0,012%	0,857%	14,8%	35,9%
<i>Newspaper</i>	0,000%	0,026%	0,043%	0,014%	0,880%	15,5%	34,4%
<i>GT_Fly</i>	0,000%	0,905%	0,865%	0,205%	1,718%	19,1%	47,2%
<i>Poznan_Hall2</i>	0,000%	0,058%	0,151%	0,051%	5,274%	17,0%	41,2%
<i>Poznan_Street</i>	0,000%	0,387%	0,503%	0,119%	1,109%	18,9%	42,3%
<i>Undo_Dancer</i>	0,000%	3,082%	3,013%	0,787%	6,253%	18,3%	47,9%
<i>Média 1024x768</i>	0,000%	0,027%	0,027%	0,011%	0,813%	15,1%	37,2%
<i>Média 1920x1088</i>	0,000%	1,108%	1,133%	0,290%	3,588%	18,3%	44,6%
<i>Média</i>	0,000%	0,645%	0,659%	0,170%	2,399%	17,0%	41,1%

1920x1088, este percentual atinge 42,4%. Avaliando-se a redução do tempo total de codificação, conclui-se através da Tabela 7 que a heurística proposta, utilizando THR 90%, atinge um ganho médio de 15,6% neste quesito. Estes dados demonstram que, embora a heurística ao utilizar THR 90% apresente um impacto considerável à eficiência de codificação de vistas sintetizadas, este proporciona uma redução de complexidade expressiva ao codificador 3D-HEVC.

Por fim, os resultados de simulação (eficiência de codificação e redução de tempo) da heurística *Early Skip/DIS* utilizando THR 80% são apresentados na Tabela 8.

Observa-se que a heurística utilizando THR 80% obteve na média um acréscimo de 0,170% em BD-rate para o canal de textura. É importante destacar que este valor médio é alavancado devido à sequência *Undo\_Dancer*, a qual apresentou um acréscimo de 0,787%. Além disso, destaca-se que a média do acréscimo em BD-rate para os quadros de textura das sequências de resolução 1024x768 é de apenas 0,011%. Assim, tendo estes dados em vista, conclui-se que o *Early Skip/DIS* ainda apresenta um impacto baixo à eficiência de codificação de quadros de textura, mesmo utilizando um THR baixo como 80%.

Percebe-se através da Tabela 8, que a eficiência de codificação de vistas sintetizadas sofre um acréscimo mais elevado ao utilizar a heurística *Early Skip/DIS* com THR 80% (+2,399% em BD-rate). Entretanto, ao observar-se apenas os dados das sequências de resolução 1024x768, verifica-se que este impacto percebido pela codificação de vistas sintetizadas é de apenas 0,813%.

Por fim, se por um lado o *Early Skip/DIS* utilizando THR 80% proporciona um acréscimo de 2,399% em BD-rate para vistas sintetizadas, por outro lado este reduz

Tabela 9: Comparação entre os resultados obtidos com a heurística *early Skip/DIS* utilizando THR 99% e 95%

THR	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Profundidade)
99%	0,012%	0,062%	9,6%	24,4%
95%	0,047%	0,409%	13,8%	33,7%
90%	0,093%	1,221%	15,6%	37,8%
80%	0,170%	2,399%	17,0%	41,1%

em média 41,1% o tempo de codificação de mapas de profundidade. Considerando apenas a sequência *Undo\_Dancer*, verifica-se que esta heurística reduz em 47,9% o tempo de codificação de mapas de profundidade, demonstrando a eficácia desta solução. Ao observar os dados de tempo total de codificação, verifica-se que, em média, a heurística proposta utilizando THR 80% apresenta ganhos de 17%, chegando a 19,1% na sequência *GT\_Fly*.

A Tabela 9 apresenta uma comparação entre os resultados médios de eficiência de codificação e redução de complexidade da solução *Early Skip/DIS* considerando os valores de THR 99%, 95%, 90% e 80%.

Nota-se que a utilização do valor de THR 99% resultou em impactos menos significativos na eficiência de codificação dos quadros de textura, entretanto apresentando resultados menos expressivos em termos de redução de complexidade, como esperado. Analogamente, a heurística *Early Skip/DIS* utilizando THR 80% proporcionou a maior redução média de tempo de codificação de mapas de profundidade (41,1%); entretanto, o acréscimo médio em BD-rate percebido na codificação de mapas de profundidade foi cerca de 2,399% neste cenário. Visando aprimorar a observação dos dados demonstrados na Tabela 9, a Figura 32 apresenta um gráfico onde são comparados os resultados redução de tempo de codificação (a) e impacto à eficiência de codificação (b) obtidos através da heurística *Early Skip/DIS* para os quatro valores de THR utilizados nos experimentos.

## 5.4 Trabalhos Relacionados e Comparações

Alguns trabalhos relacionados à redução de complexidade do codificador 3D-HEVC foram encontrados na literatura. O trabalho desenvolvido em (TOHIDYPOUR; POURAZAD; NASIOPOULOS, 2015) apresenta um esquema de redução de complexidade para o processo de codificação de quadros de textura do 3D-HEVC. Para tanto, utiliza modelos probabilísticos visando predizer qual o modo com o menor custo RD das CUs filhas de uma dada CU raiz.

Os autores de (SANCHEZ et al., 2014a) propõem um filtro denominado GMOF

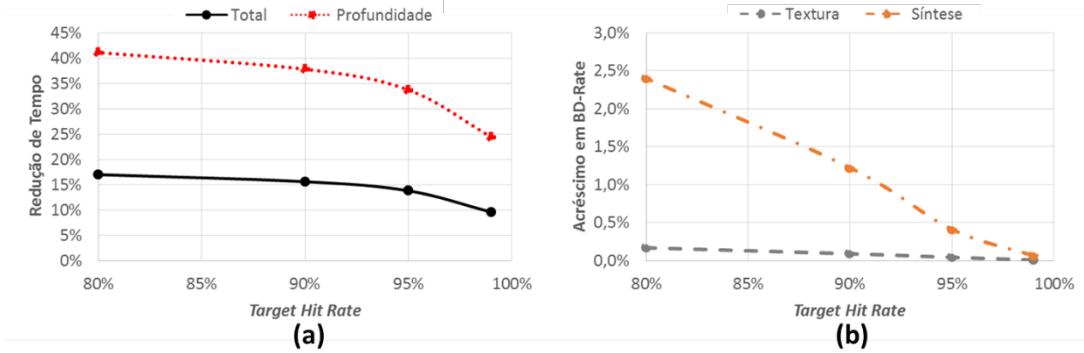


Figura 32: Gráfico comparativo entre resultados de redução de tempo (a) e perda na eficiência de codificação (b) obtidos através da heurística *Early Skip/DIS* utilizando THR 99%, 95%, 90% e 80%.

(*Gradient-Based Mode One Filter*), que visa reduzir a complexidade da aplicação do modo DMM do 3D-HEVC. Este filtro visa reduzir o número de *Wedgelets* avaliadas durante o processo de teste do modo DMM.

Um algoritmo simplificado detector de bordas é descrito em (SANCHEZ et al., 2014b), o qual permite classificar os blocos de codificação em duas categorias: bordas ou regiões praticamente constantes. Desta forma, os autores propõem evitar o teste de determinados modos intraquadro que apresentariam uma baixa probabilidade de ocorrer de acordo com as amostras de profundidade compreendidas pelo bloco de codificação.

O trabalho descrito em (PARK, 2015) apresenta um algoritmo de decisão de modo rápido, que é baseado em classificação de bordas utilizando a transformada Hadamard. O esquema apresentado por Park (PARK, 2015) foca na redução de complexidade da codificação de mapas de profundidade do 3D-HEVC.

Por fim, os autores de (ZHANG et al., 2013) propõem um algoritmo de redução do número de modos intraquadro convencionais (estipulados pelo HEVC) e DMMs avaliados durante o processo de codificação de mapas de profundidade, de forma adaptativa. Para tanto, a solução proposta explora características da imagem de profundidade e dos *pixels* de referência da predição intra.

A Tabela 10 apresenta uma comparação entre os resultados obtidos com a heurística *Early Skip/DIS* com os trabalhos relacionados. O trabalho descrito em (TOHIDYPOUR; POURAZAD; NASIOPOULOS, 2015) apresenta estratégias de redução de complexidade para o codificador de textura do 3D-HEVC, por isto, o artigo não apresenta o dado de redução de complexidade do codificador de mapas de profundidade. Além disso, destaca-se que apenas os resultados utilizando THR 99% e 90% foram utilizados nas comparações, visto que os ganhos em termos de redução de tempo de codificação de mapas de profundidade obtidos com THR 95% foram suficientes para superar os demais trabalhos relacionados.

Tabela 10: Comparação da heurística *Early Skip/DIS* com trabalhos relacionados

Trabalho	Apenas Textura (BD-rate)	Apenas Síntese (BD-rate)	Red. de Tempo (Total)	Red. de Tempo (Prof.)
(TOHIDYPOUR et al., 2015)	0,350%	0,310%	23,0%	-
(SANCHEZ et al., 2014a)	0,022%	-0,047%	2,7%	9,8%
(SANCHEZ et al., 2014b)	0,027%	-0,064%	5,9%	23,8%
(PARK, 2015)	-0,030%	0,130%	-	22,2%
(ZHANG et al., 2013)	-	1,030%	-	27,9%
<b>THR 99%</b>	<b>0,012%</b>	<b>0,062%</b>	<b>9,6%</b>	<b>24,4%</b>
<b>THR 95%</b>	<b>0,047%</b>	<b>0,409%</b>	<b>13,8%</b>	<b>33,7%</b>

Considerando THR 99%, observa-se que, dentre os trabalhos relacionados, a heurística apresenta um dos menores impactos em termos de eficiência de codificação, tanto na avaliação do BD-rate para os quadros de textura quanto para as vistas sintetizadas. Além disso, comparado com os trabalhos relacionados que apresentaram dados de redução de complexidade total, observa-se que a heurística proposta utilizando THR 99% obteve o segundo melhor resultado neste quesito.

Ao comparar apenas com os trabalhos que propõem soluções para o codificador de mapas de profundidade, percebe-se que, embora a heurística Early Skip/DIS utilizando THR 95% apresente o maior acréscimo em BD-rate na codificação de quadros de textura e o segundo maior no de vistas sintetizadas, essa também proporciona a maior redução média de tempo de codificação de mapas de profundidade. Assim, demonstrando relevância da heurística proposta nesta dissertação.

## 5.5 Considerações Finais do Capítulo

Este capítulo apresentou a heurística *Early Skip/DIS* desenvolvida, a qual é baseada na avaliação dos custos RD dos modos *Skip* e *DIS* e na geração de *thresholds* adaptativos. Essa heurística evita a avaliação dos diversos modos e ferramentas de codificação do 3D-HEVC com base no custo RD gerado pelos modos *Skip* e *DIS*. Uma análise preliminar, observando taxa de ocorrência dos modos *Skip* e *DIS* apontou que a probabilidade de ocorrência destes modos é inversamente proporcional ao custo RD gerado. Além disto, esta análise também demonstrou que diferentes sequências de vídeo apresentam curvas de ocorrência distintas, gerando a necessidade do uso de *thresholds* adaptativos. Por fim, a análise também demonstrou que o tamanho do bloco influencia na taxa de ocorrência dos modos avaliados, apontando a necessidade de se gerar *thresholds* distintos para cada tamanho de bloco.

Resultados de simulação demonstraram que a heurística *Early Skip/DIS* proposta,

utilizando uma taxa de acerto alvo THR de 99%, é capaz de reduzir a complexidade do processo de codificação de mapas de profundidade em 24,4%, em média. O impacto percebido na codificação de textura neste cenário é de apenas 0,012% em termos de BD-rate. Além disso, a codificação das vistas sintetizadas experimentou uma redução na eficiência de codificação (acréscimo em BD-rate) de 0,062%. Considerando THR 95%, a heurística *Early Skip/DIS* é capaz de reduzir em um terço (33,7% em média) o tempo de codificação dos mapas de profundidade, ao preço de apenas 0,047% de acréscimo no BD-rate para o canal de textura e 0,409% para as vistas sintetizadas.

Valores de THR ainda mais inferiores proporcionam uma redução de complexidade um pouco superior às obtidas com 99% e 95%. Entretanto, o acréscimo em BD-rate para vistas sintetizadas atinge, em média, 1,221% e 2,399% utilizando os THRs 90% e 80%, respectivamente. Isto demonstra que a heurística *Early Skip/DIS*, utilizando valores mais inferiores de THR, prejudica a eficiência de codificação das vistas sintetizadas do 3D-HEVC.

É importante destacar que a solução heurística proposta nesta dissertação foi testada em sequências de vídeo que possuem, em sua grande maioria, 10 segundos de duração e sem mudanças abruptas nas cenas. Além disso, destaca-se que as características de um vídeo tendem a variar muito de uma cena para outra. Portanto, observa-se que ao incorporar esta heurística em uma aplicação real, será necessário atualizar periodicamente os *thresholds* do algoritmo *Early Skip/DIS*.

Comparado com trabalhos relacionados, a heurística *Early Skip/DIS* apresentou resultados competitivos em termos de degradação na eficiência de codificação das vistas de textura e sintetizadas. Considerando o tempo total de codificação, a heurística, utilizando THRs 99% e 95%, apresentou o segundo melhor resultado de redução total de complexidade, perdendo apenas para uma solução que também reduz a complexidade no processamento de quadros de textura. Avaliando apenas a redução de complexidade para o codificador de mapas de profundidade, a heurística utilizando THR 95% apresentou os melhores resultados neste quesito entre todos os trabalhos relacionados.

Desta forma, conclui-se que a heurística *Early Skip/DIS*, utilizando maiores valores de THR, possui uma contribuição importante para a área de codificação de vídeos, atingindo resultados expressivos em termos de redução de complexidade na codificação dos mapas de profundidade, com impactos mínimos na eficiência de codificação do 3D-HEVC.

## 6 CONCLUSÃO

Esta dissertação apresentou uma heurística de redução de complexidade para o 3D-HEVC, chamada *Early Skip/DIS* a qual visa reduzir o tempo de codificação dos mapas de profundidade, a fim de apresentar um impacto reduzido na eficiência de codificação do padrão.

Uma análise de distribuição dos modos de particionamento apresentada no Capítulo 4, possibilitou observar a ocorrência extremamente baixa de PUs assimétricas e retangulares no processo de codificação de mapas de profundidade no 3D-HEVC e uma grande concentração de uso do modo de partição 2Nx2N, chegando próximo a 100% em alguns casos. Este fato conduziu à análise da remoção dos modos menos usados no processo de codificação. Resultados de simulação demonstraram que a remoção de todos os modos assimétricos e retangulares acarretou em uma redução de 6,1% no tempo de codificação dos mapas de profundidade, apresentando um aumento de apenas 0,001% em BD-rate no canal de textura e 0,084% nas vistas sintetizadas. Com isto demonstrou-se que, além de pouco utilizados, estes modos também trazem impacto muito baixo tanto na eficiência de codificação quanto no tempo de execução.

O uso elevado da partição 2Nx2N e sua relevância na codificação de mapas de profundidade é função direta das ferramentas *Skip*, *DIS* e *Merge*, que são usadas apenas com o modo de particionamento 2Nx2N. Conforme estudos demonstrados no Capítulo 5, tais ferramentas são usadas em mais de 90% dos casos avaliados e são capazes de gerar uma predição qualificada das amostras do vídeo utilizando pouca informação, conduzindo a um custo RD atraente. Este fato, além de justificar a baixa ocorrência dos modos de particionamento assimétricos e retangulares, conduziu a decisão de desenvolver a heurística de decisão antecipada apresentada no Capítulo 5, denominada *Early Skip/DIS*, que é a principal contribuição deste trabalho.

A heurística *Early Skip/DIS* avalia, primeiramente, o modo *Skip* e, caso o custo RD gerado seja menor que um determinado *threshold*, os demais modos não são avaliados. Caso contrário, o modo *DIS* é avaliado e o menor custo RD entre os custos gerados pelos modos *Skip* e *DIS* é comparado com um segundo *threshold*. Se o custo

RD obtido for menor que o segundo *threshold*, o modo que apresentou melhor custo de codificação é selecionado para o bloco que está sendo processado, caso contrário, os demais modos são avaliados.

Os *thresholds* utilizados na heurística *Early Skip/DIS* são definidos de forma adaptativa, observando a taxa de ocorrência dos modos. Além disso, destaca-se que a solução prioriza a seleção do modo *Skip* em relação ao modo *DIS*, devido à taxa de ocorrência e aos ganhos de eficiência de codificação proporcionados por esta ferramenta.

Resultados de simulação demonstraram que a heurística, utilizando uma taxa de acerto alvo de 99%, reduziu a complexidade do processo de codificação de mapas de profundidade em 24,4%. Considerando a taxa de acerto alvo de 9%, a heurística *Early Skip/DIS* apresentou uma redução de um terço do tempo de codificação dos mapas de profundidade. Em ambos os cenários, a eficiência de compressão dos quadros de textura não sofre impactos significativos, percebendo um acréscimo em BD-rate inferior a 0,2% no pior caso avaliado. Além disso, a eficiência de codificação das vistas sintetizadas sofre impactos médios de 0,062% e 0,409% utilizando THR 99% e 95% respectivamente. Avaliou-se também o uso de THRs ainda mais inferiores (90% e 80%), a redução média de complexidade para o codificador de mapas de profundidade foi de 37,8% a 41,1%. Apesar destes expressivos resultados de redução de tempo de codificação, o impacto percebido pelo codificador de mapas de profundidade varia entre 1,222% a 2,399% em BD-rate na média. Assim, conclui-se que o ponto de operação ideal do *Early Skip/DIS* está entre os valores 99% e 95% de THR, onde atinge-se resultados expressivos em termos de redução de complexidade ao preço de um baixo impacto na eficiência de codificação.

Comparado com trabalhos relacionados, a heurística proposta atingiu os melhores resultados em termos de redução de tempo de codificação dos mapas de profundidade. Além disso, os resultados obtidos considerando o impacto na eficiência de codificação do canal de textura está equivalente aos demais trabalhos. Considerando os THRs 99% e 95%, a eficiência do codificador de textura e de vistas sintetizadas, observa-se que a solução *Early Skip/DIS* apresenta resultados competitivos com os demais trabalhos. É importante destacar que os trabalhos relacionados foram publicados nos principais eventos e revistas de codificação, processamento e tecnologia de imagens e vídeos digitais. Assim, é possível perceber a relevância da contribuição da heurística apresentada neste trabalho.

Como trabalhos futuros, pretende-se desenvolver uma solução de decisão antecipada de modo para o codificador dos quadros de textura do 3D-HEVC. Além disso, espera-se explorar outras estratégias de redução de complexidade para o codificador de textura e mapas de profundidade do 3D-HEVC. Tais estratégias, assim como as propostas neste trabalho, visarão reduzir o tempo de codificação sem impactar signifi-

cativamente a eficiência de codificação.

## REFERÊNCIAS

- AGOSTINI, L. V. **Desenvolvimento de Arquiteturas de Alto Desempenho Dedicadas à Compressão de Vídeo Segundo o Padrão H.264/AVC.** 2007. Tese (Doutorado em Ciência da Computação) — Universidade Federal do Rio Grande do Sul.
- BJONTEGAARD, G. **Calculation of average PSNR differences between RD-curves.** [S.I.]: ITU-T, 2001.
- BOSSEN, F.; BROSS, B.; SUHRING, K.; FLYNN, D. HEVC Complexity and Implementation Analysis. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.I.], v.22, n.12, p.1685–1696, Dec 2012.
- CHEN, Y.; TECH, G.; WEGNER, K.; YEA, S. **Test Model 11 of 3D-HEVC and MV-HEVC.** [S.I.]: Joint Collaborative Team on 3D Video Coding Extension, 2015.
- CORREA, G.; ASSUNCAO, P.; AGOSTINI, L.; SILVA CRUZ, L. da. Performance and Computational Complexity Assessment of High-Efficiency Video Encoders. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.I.], v.22, n.12, p.1899–1909, Dec 2012.
- FEHN, C.; ATZPADIN, N.; MULLER, M.; SCHREER, O.; SMOLIC, A.; TANGER, R.; KAUFF, P. An Advanced 3DTV Concept Providing Interoperability and Scalability for a Wide Range of Multi-Baseline Geometries. In: **IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING**, 2006., 2006. **Anais...** [S.I.: s.n.], 2006. p.2961–2964.
- GHANBARI, M. **Standard Codecs:** Image Compression to Advanced Video Coding. [S.I.]: Institution Electrical Engineers, 2003.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing, Third Edition.** [S.I.]: Prentice Hall, 2008.
- ITU-T. **High efficiency video coding:** Recommendation ITU-T H.265. [S.I.]: International Telecommunication Union: Telecommunication Standardization Sector of ITU, 2014.

JCT-VC. **The First JCT-VC Meeting**. Dresden: International Telecommunication Union ITU-T, 2010.

JVT. **Joint Multiview Video Coding**. [S.I.: s.n.], 2009.

KAUFF, P.; ATZPADIN, N.; FEHN, C.; MÄLLER, M.; SCHREER, O.; SMOLIC, A.; TANGER, R. Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. **Signal Processing: Image Communication**, [S.I.], v.22, n.2, p.217 – 234, 2007. Special issue on three-dimensional video and television.

LATHI, B. **Modern Digital and Analog Communications Systems**. [S.I.]: Oxford University Press, Incorporated, 1998. (The Oxford Series in Electrical and Computer Engineering).

LEE, J. Y.; PARK, M. W.; KIM, C. **3D-CE1**: Depth intra skip (DIS) mode. [S.I.]: Joint Collaborative Team on 3D Video Coding Extension, 2015.

MCCANN, K.; BROSS, B.; HAN, W.-J.; KIM, I. K.; SUGIMOTO, K.; SULLIVAN, G. J. **High Efficiency Video Coding (HEVC) Test Model 8 (HM 8) Encoder Description**. [S.I.]: Joint Collaborative Team on Video Coding, 2012.

MULLER, K.; SCHWARZ, H.; MARPE, D.; BARTNIK, C.; BOSSE, S.; BRUST, H.; HINZ, T.; LAKSHMAN, H.; MERKLE, P.; RHEE, F.; TECH, G.; WINKEN, M.; WIEGAND, T. 3D High-Efficiency Video Coding for Multi-View Video and Depth Data. **Image Processing, IEEE Transactions on**, [S.I.], v.22, n.9, p.3366–3378, Sept 2013.

PARK, C.-S. Edge-Based Intramode Selection for Depth-Map Coding in 3D-HEVC. **Image Processing, IEEE Transactions on**, [S.I.], v.24, n.1, p.155–162, Jan 2015.

RICHARDSON, I. E. **Video Codec Design**: Developing Image and Video Compression Systems. New York, NY, USA: John Wiley & Sons, Inc., 2002.

RICHARDSON, I. E. G. **H.264 and MPEG-4 Video Compression**: Video Coding for Next-generation Multimedia. New York, NY, USA: John Wiley & Sons, Inc., 2003.

ROSEWARNE, C.; BROSS, B.; NACCARI, M.; SHARMAN, K.; SULLIVAN, G. **High Efficiency Video Coding (HEVC) Test Model 16 (HM 16)**: Improved Encoder Description. [S.I.]: Joint Collaborative Team on Video Coding, 2015.

RUSANOVSKY, D.; MULLER, K.; VETRO, A. **Common Test Conditions of 3DV Core Experiments**. [S.I.]: Joint Collaborative Team on 3D Video Coding Extension, 2013.

- SALOMON, D. **Data Compression**: The Complete Reference. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- SANCHEZ, G.; SALDANHA, M.; BALOTA, G.; ZATT, B.; PORTO, M.; AGOSTINI, L. A complexity reduction algorithm for depth maps intra prediction on the 3D-HEVC. In: VISUAL COMMUNICATIONS AND IMAGE PROCESSING CONFERENCE, 2014 IEEE, 2014. *Anais...* [S.I.: s.n.], 2014. p.137–140.
- SANCHEZ, G.; SALDANHA, M.; BALOTA, G.; ZATT, B.; PORTO, M.; AGOSTINI, L. Complexity reduction for 3D-HEVC depth maps intra-frame prediction using simplified edge detector algorithm. In: IMAGE PROCESSING (ICIP), 2014 IEEE INTERNATIONAL CONFERENCE ON, 2014. *Anais...* [S.I.: s.n.], 2014. p.3209–3213.
- SHI, Y. Q.; SUN, H. **Image and Video Compression for Multimedia Engineering**: Fundamentals, Algorithms, and Standards. 2nd.ed. Boca Raton, FL, USA: CRC Press, Inc., 2008.
- SULLIVAN, G.; OHM, J.; HAN, W.-J.; WIEGAND, T. Overview of the High Efficiency Video Coding (HEVC) Standard. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.I.], v.22, n.12, p.1649–1668, Dec 2012.
- SPRINGER (Ed.). **High Efficiency Video Coding (HEVC)**: Algorithms and Architectures. [S.I.]: Springer International Publishing Switzerland, 2014.
- TECH, G.; CHEN, Y.; MULLER, K.; OHM, J.-R.; VETRO, A.; WANG, Y.-K. Overview of the Multiview and 3D Extensions of High Efficiency Video Coding. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.I.], v.PP, n.99, p.1–1, 2015.
- TOHIDYPOUR, H.; POURAZAD, M.; NASIOPOULOS, P. Online Learning-based Complexity Reduction Scheme for 3D-HEVC. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.I.], v.PP, n.99, p.1–1, 2015.
- WIEGAND, T.; SULLIVAN, G.; BJONTEGAARD, G.; LUTHRA, A. Overview of the H.264/AVC video coding standard. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.I.], v.13, n.7, p.560–576, July 2003.
- ZHANG, N.; CHEN, Y.-W.; LIN, J.-L.; FAN, X.; MA, S.; ZHAO, D.; GAO, W. Improved disparity vector derivation in 3D-HEVC. In: VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), 2013, 2013. *Anais...* [S.I.: s.n.], 2013. p.1–5.

## APÊNDICE A FERRAMENTA 3D-HEVC ANALYZER

Este apêndice descreve, resumidamente, uma ferramenta desenvolvida ao longo do mestrado que permite analisar a taxa de ocorrência de determinados modos e ferramentas de codificação, tendo em vista um determinado cenário. Devido à sua função, esta ferramenta foi nomeada *3D-HEVC Analyzer*. Destaca-se que as análises apresentadas na seção 4.1 foram realizadas com o auxílio desta ferramenta.

O *3D-HEVC Analyzer* foi desenvolvido na linguagem de programação C#, utilizando o ambiente de desenvolvimento integrado (*Integrated Development Environment* - IDE) *Microsoft Visual Studio*.

A Figura 33 apresenta a interface da ferramenta *3D-HEVC Analyzer*. Dois arquivos no formato *csv* (*Comma-Separated Values*) compõem a entrada da ferramenta, o *Compressing Evaluation* e o *Compressed Evaluation*, os quais são gerados pelo software de referência modificado. O primeiro contém informações sobre informações que podem ser obtidas durante o processo de codificação, tal como: o tipo de *Slice* (SUL-LIVAN et al., 2012), se o quadro que está sendo codificado é textura ou mapa de profundidade, o tamanho da CU, média das amostras, etc.

O segundo arquivo contém informações sobre as tomadas de decisão do codificador para cada CU, tal como o modo de predição (*Intra* ou *Inter*), o modo de particionamento, se a CU se subdividiu em outras 4 CUs filhas, etc. Além disso, este arquivo também contém informações se a CU utilizou as ferramentas *Merge*, *Skip* e/ou *DIS*.

Os arquivos *csv* utilizados pela ferramenta são gerados através de modificações realizadas no software de referência do 3D-HEVC. Desta forma, ao codificar uma sequência de vídeo com o HTM modificado, o mesmo gerará os arquivos *compressingEvaluation.csv* e *compressedEvaluation.csv*. A visão geral do sistema utilizando a ferramenta *3D-HEVC Analyzer* é ilustrada na Figura 34. Observa-se que a única diferença percebida no processo de codificação é a geração dos arquivos *csv*, não havendo a necessidade de outras informações de entrada, ou modificações no *bitstream* gerado.

Uma vez carregados os arquivos *csv* na ferramenta, é possível selecionar em *Compressing Information* qual o contexto de codificação que se deseja avaliar a taxa

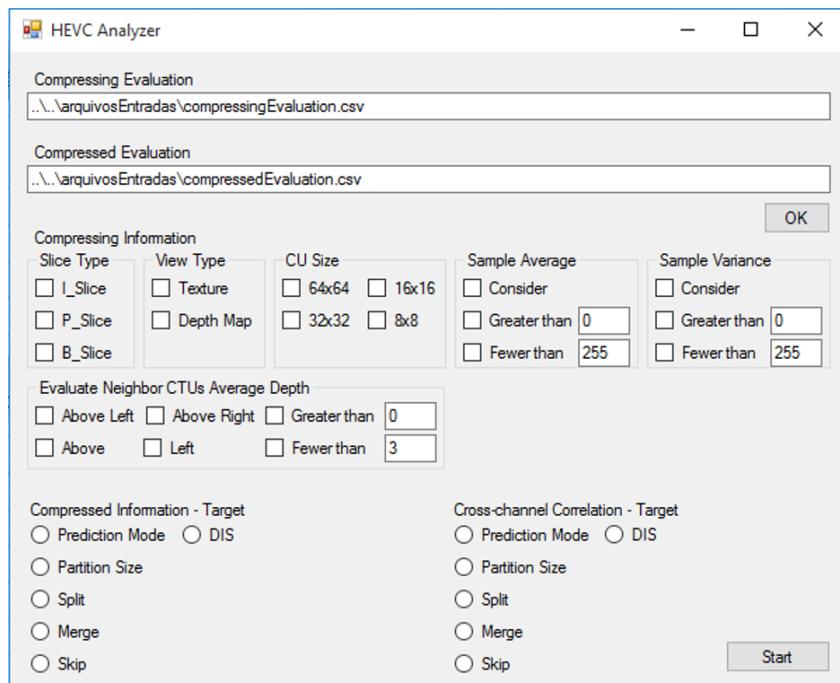


Figura 33: Interface de usuário da ferramenta *3D-HEVC Analyzer*

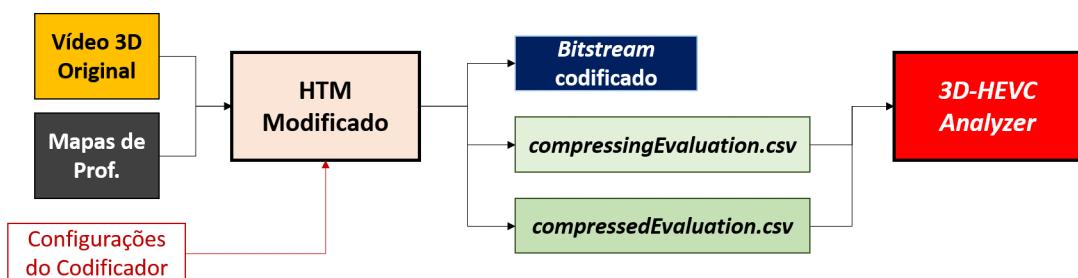


Figura 34: Visão geral do sistema utilizando a ferramenta *3D-HEVC Analyzer*

de ocorrência dos modos e das ferramentas de codificação. A Figura 35 apresenta um exemplo de uso do *3D-HEVC Analyzer*, onde avalia-se apenas CUs de tamanho 64x64 e 32x32 pertencentes a mapas de profundidade, cuja média das amostras que compõem o bloco seja inferior a 128. Variância e Profundidade da QT das CTUs vizinhas não são consideradas neste exemplo. Além disso, todos os tipos de *Slices* são avaliados.

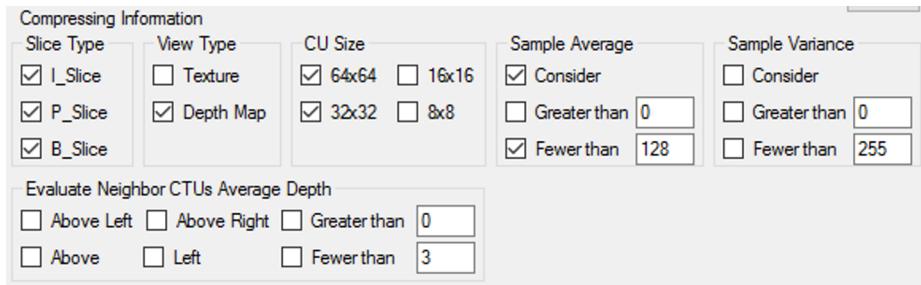


Figura 35: Exemplo de seleção para os itens em *Compressing Information* na ferramenta *3D-HEVC Analyzer*

Uma vez definido o contexto de codificação que deseja-se avaliar, é possível determinar a taxa de ocorrência de qual modo ou ferramenta deseja-se analisar. Além disso, a ferramenta também permite avaliar a taxa de ocorrência utilizando informações cruzadas entre os canais (textura ↔ profundidade), ou seja, fazer uma análise considerando a tomada de decisão da CU atual e da CU co-localizada no outro canal.

A Figura 36 demonstra um exemplo de seleção na ferramenta, que visa avaliar a taxa de ocorrência do modo *Skip*, dado o cenário especificado em *Compressing Information*.

Ao pressionar *Start*, o *3D-HEVC Analyzer* inicializa o processo de análise de ocorrência, conforme as especificações dadas. Uma vez finalizado o processo de análise, o resultado é exibido ao usuário. Um exemplo de resultado utilizando a seleção efetuada na Figura 36 é demonstrada na Figura 37.

Observa-se na Figura 37 que o número total de CUs analisadas é fornecido ao usuário (37.871). Além disso, o programa informa o total de CUs inválidas avaliadas durante a análise. Uma CU é considerada inválida quando esta é avaliada durante o processo de compressão (*Compressing*), entretanto o codificador determinou para aquela região uma CU de tamanho maior ou menor. Desta forma, considerando as especificações realizadas em *Compressing Information* na Figura 36, uma CU pertencente ao conjunto de CUs inválidas “*CompressingSize < CompressedSize*” significa uma CU de tamanho 32x32 avaliada durante o processo de codificação, cujo tamanho definido para aquela região tenha sido uma CU de tamanho 64x64.

Por fim, na Figura 37 é demonstrada a quantidade de CUs codificadas (ou não) com o modo *Skip*. No exemplo apresentado, observa-se que 72,33% destas selecionaram o *Skip*. Isto considerando apenas CUs classificadas como válidas.

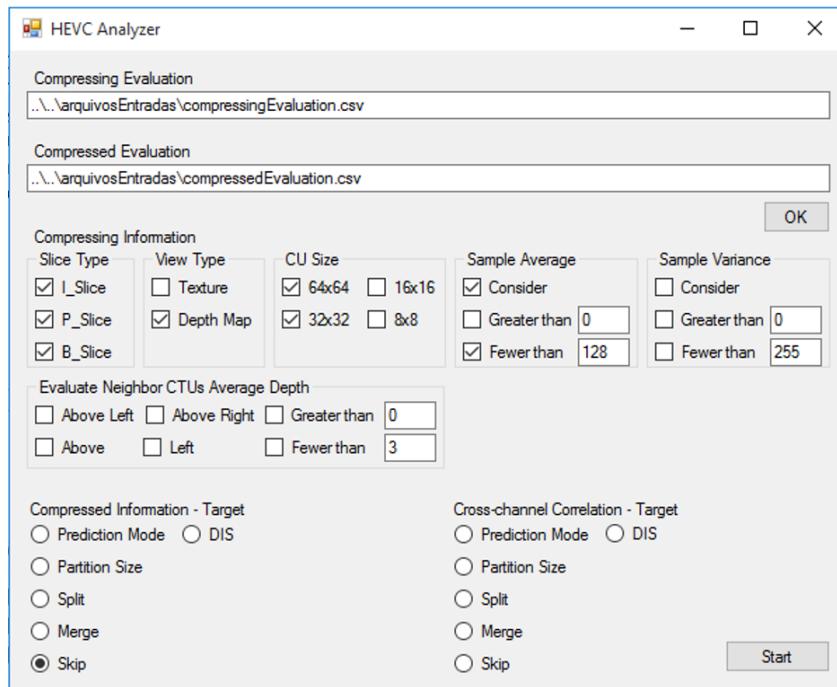


Figura 36: Exemplo de uso da ferramenta 3D-HEVC Analyzer

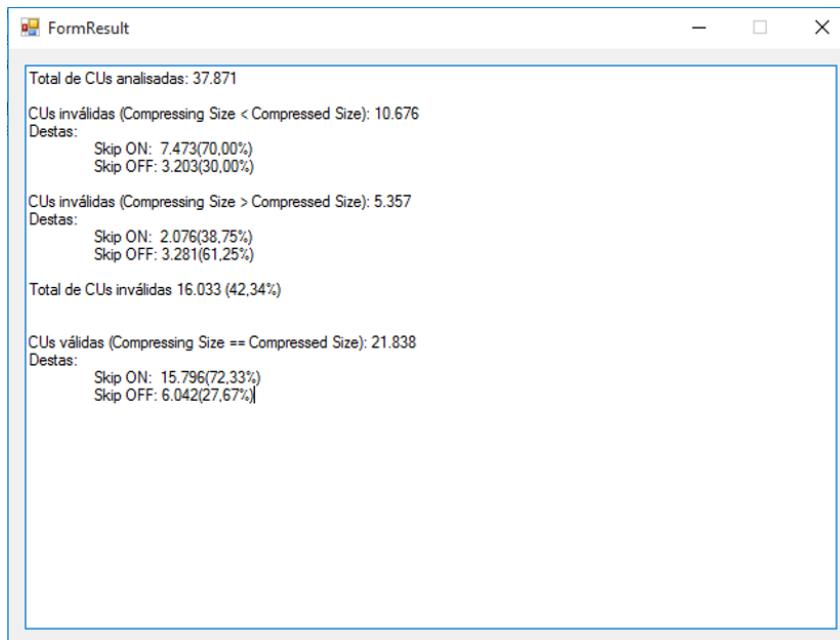


Figura 37: Exemplo de resultado fornecido pela ferramenta 3D-HEVC Analyzer

A Figura 38 demonstra um exemplo de saída para a ferramenta *3D-HEVC Analyzer*, observando a ocorrência do modo *Skip* utilizando informações cruzadas entre os canais de textura e profundidade. Considerando as especificações realizadas em *Compressing Information*, na Figura 35, *sedCU* e *crossCU* significam, respectivamente, a CU do canal de profundidade (canal atual) e a CU do canal de profundidade (canal cruzado).

Segundo o exemplo demonstrado na Figura 38, 16.296 CUs co-localizadas de textura e profundidade tomaram a mesma decisão em relação ao modo *Skip*. Destas, 81,81% foram codificadas com o mesmo tamanho (64x64 ou 32x32).

Somando-se a estas informações, matrizes de análise cruzada também são fornecidas ao usuário. Estas demonstram quantas vezes a CU do canal atual tomou uma determinada decisão, comparando com a decisão do canal cruzado. Exemplificando, considerando a segunda matriz apresentada no Figura 38, observa-se que em 1.425 casos a CU do canal atual decidiu utilizar o *Skip* (ON) e o canal cruzado não (OFF), ao passo que em 11.679 casos ambas CUs decidiram utilizar o modo *Skip*.

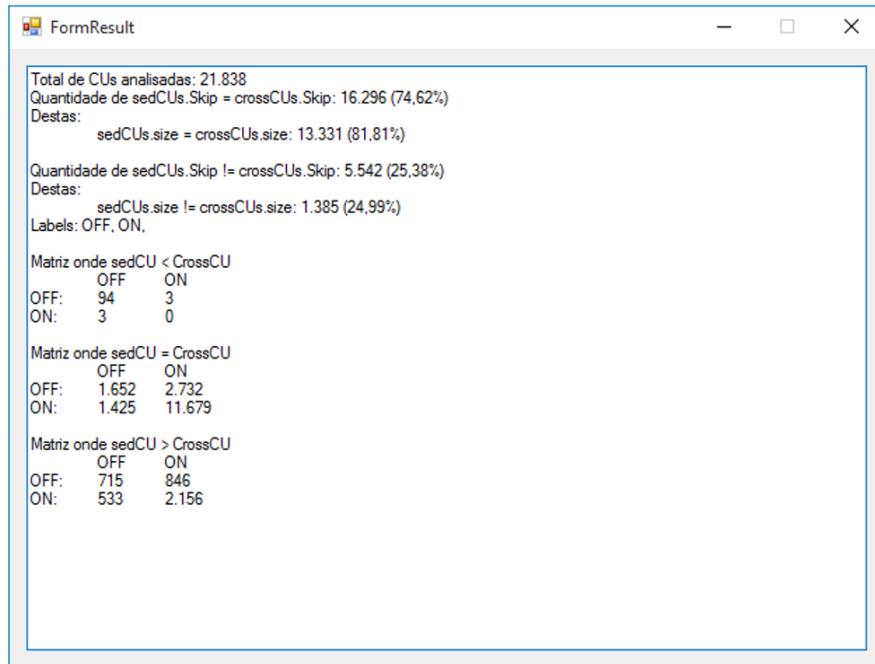


Figura 38: Exemplo de resultado fornecido pela ferramenta *3D-HEVC Analyzer* utilizando informações cruzadas entre os canais

## APÊNDICE B TRACER HEVC/3D-HEVC

Este Apêndice apresenta o *Tracer* para o HEVC/3D-HEVC que foi desenvolvido no escopo deste trabalho, mais especificamente no período de estudo e avaliação do software de referência do HEVC e 3D-HEVC. Esta ferramenta plota os particionamentos de CUs e PUs, tal como o tipo de predição utilizado em cada CU, pelo módulo da decisão de modo em cada quadro codificado, viabilizando, assim uma percepção visual de quais modos ocorrem frequentemente, e em quais cenários. A ferramenta *Tracer* foi desenvolvida em *Matlab*, utilizando seu respectivo ambiente de desenvolvimento.

A Figura 39 apresenta um diagrama de blocos do sistema utilizado para gerar os traces de codificação através da ferramenta desenvolvida. No caso em que se está processando uma sequência 2D, o vídeo original é codificado utilizando o software de referência do HEVC, o *HEVC Model* (HM), com algumas modificações, visando gerar o arquivo *trace.csv*, o qual armazena as informações de tomada de decisão do codificador. Quando processando uma sequência 3D, o vídeo original é codificado utilizando o software de referência do 3D-HEVC, o *3D-HEVC Test Model* (HTM), com as mesmas modificações inseridas no HM.

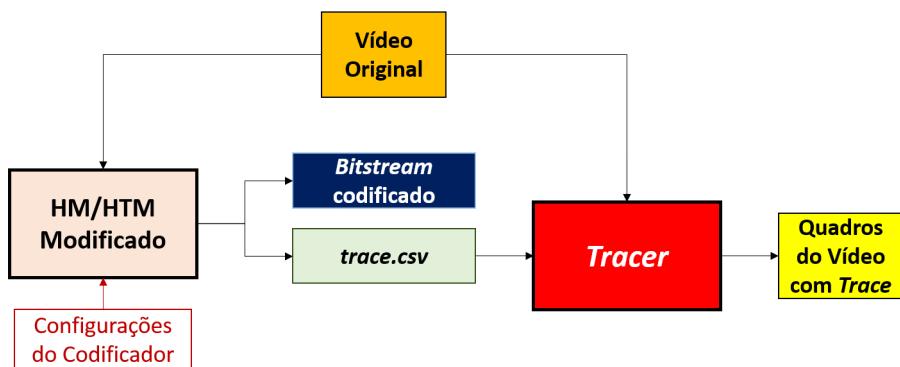


Figura 39: Sistema para geração dos traces de codificação

O arquivo *trace.csv* gerado pelo HM/HTM é utilizado como entrada da ferramenta *Tracer* juntamente com o vídeo original, o qual foi processado pelo codificador. Com isto, para cada quadro codificado, o *Tracer* gera os quadros do vídeo em formato de

imagem PNG (*Portable Network Graphics*), com as tomadas de decisão destacas nos próprios quadros.

A Figura 40 demonstra um exemplo de *trace* aplicado ao vídeo *Balloons* sem seus quadros de textura e mapa de profundidade da vista independente. Ao passar pelo *Tracer*, o vídeo é descolorido, e todas as CUs são destacadas em linhas pretas. Após, CUs que são codificadas com predição interquadros, são destacadas em azul, ao passo que CUs codificadas com predição intraquadro, são coloridas em vermelho. As CUs que empregam o modo *Skip* continuam em escala de cinza.

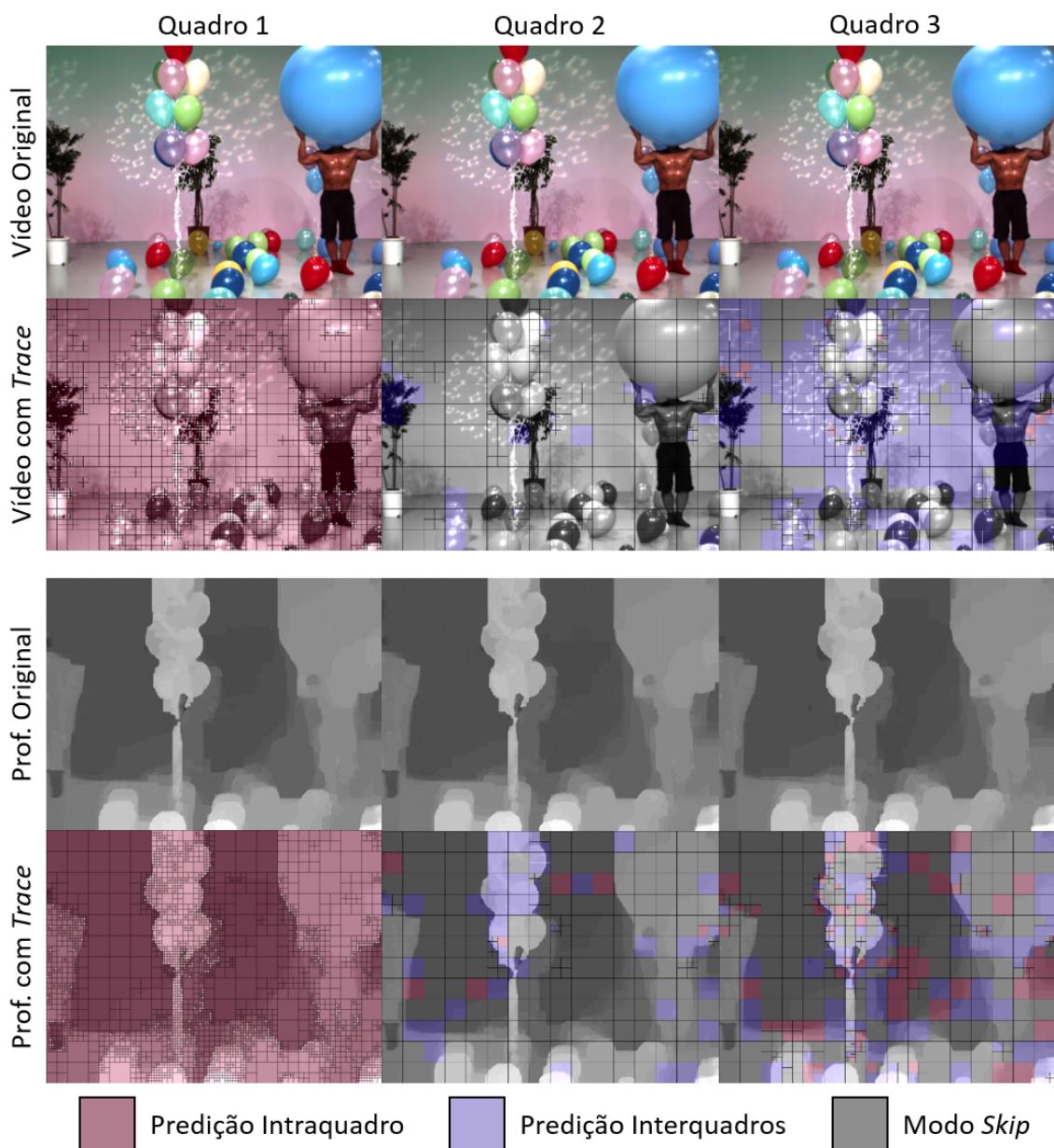


Figura 40: Textura e mapa de profundidade originais do vídeo *Balloons* mais *trace* gerado para a vista independente

É importante salientar que o *trace* gerado não apresenta todos os modos e ferramentas estipulados no HEVC/3D-HEVC, como o *Merge* e o *DIS* por exemplo. Estes, ao passarem pelo *Tracer*, são interpretados como predição interquadros e intraquadro, respectivamente.

Além de destacar modos e ferramentas de codificação utilizados, o *Tracer* também plota, na imagem gerada, o particionamento de CUs em PUs realizado pelo codificador, desenhando linhas brancas nestas regiões. Visto que estas informações não estão claramente visíveis na Figura 40, a Figura 41 apresenta o trace obtido para um quadro da sequência de vídeo *RaceHorses*. As barras pretas presentes ao lado direito e abaixo do quadro do vídeo são inseridas pelo *Tracer* visando tornar a resolução do mesmo múltipla de 64 (tamanho da CTU) em ambos eixos.

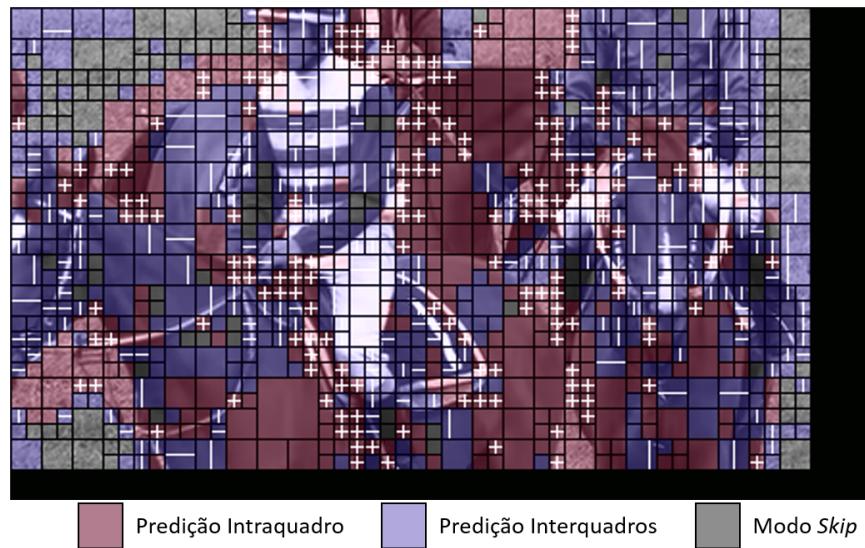


Figura 41: *Trace* obtido para um quadro da sequência de vídeo *RaceHorses*