

**UNIVERSIDADE DO EXTREMO SUL CATARINENSE - UNESC**

**CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**ALLAN SERRA BRAGA BUGYI**

**ANÁLISE COMPARATIVA ENTRE OS PADRÕES ITU-T H.265/HEVC E H.264/AVC  
EM CODIFICAÇÃO DE VÍDEO**

**CRICIÚMA**

**2015**

**ALLAN SERRA BRAGA BUGYI**

**ANÁLISE COMPARATIVA ENTRE OS PADRÕES ITU-T H.265/HEVC E H.264/AVC  
EM CODIFICAÇÃO DE VÍDEO**

Trabalho de Conclusão de Curso, apresentado para obtenção do grau de Bacharel no curso de Ciência da Computação da Universidade do Extremo Sul Catarinense, UNESC.

Orientador: Prof. MEng. Evânio Ramos Nicoleit

**CRICIÚMA**

**2015**


**ALLAN SERRA BRAGA BUGYI**

**ANÁLISE COMPARATIVA ENTRE OS PADRÕES ITU-T H.265/HEVC E  
H.264/AVC EM CODIFICAÇÃO DE VÍDEO**

Trabalho de Conclusão de Curso aprovado pela Banca Examinadora para obtenção do Grau de Bacharel, no Curso de Ciências da Computação da Universidade do Extremo Sul Catarinense, UNESC, com Linha de Pesquisa em Codificação de Vídeo.

Criciúma, 25 de novembro de 2015.

**BANCA EXAMINADORA**

  
Prof. Evânio Ramos Nicoleit – M.Eng. - UNESC - Orientador

  
Prof.ª Leila Lais Gonçalves – M.Sc. - UNESC

  
Prof. Paulo João Martins – M.Sc. - UNESC

**Dedico esta trabalho à minha família pelo  
suporte incondicional e ao meu avô  
Alexsandor Bugyi.**

## **AGRADECIMENTOS**

Agradeço, primeiramente, a Deus, sem O qual eu não seria nada, pelo dom da vida, pelas oportunidades e, pela capacidade que a mim me concedeu de poder completar mais este ciclo.

Agradeço aos meus pais Ricardo Bugyi e Rosely Serra Braga Bugyi e a minha irmã Ellen Serra Braga Bugyi, os quais sempre me apoiaram em momentos bons ou ruins, me aconselharam com sabedoria e me trataram com amor. Sem o suporte deles não teria conquistado o que conquistei. Certamente que grande parte da minha energia e motivação é baseada na minha família.

Agradeço, também, ao meu falecido avô Alexsandor Bugyi, alguém que foi muito especial na minha formação e crescimento. Sou grato pelas tantas coisas que pude aprender com sua trajetória e experiência e, pelo tempo, energia, atenção e recursos que me confiou em diferentes momentos da minha vida.

**“Não há nada que seja maior evidência de insanidade do que fazer a mesma coisa dia após dia e esperar resultados diferentes.”**

**Albert Einstein**

## RESUMO

O presente trabalho realiza análise comparativa do tipo objetiva em codificação de vídeo. São objetos de estudo os padrões criados pelas organizações, *International Telecommunication Union – Telecommunication Sector* (ITU-T) e *ISO/IEC Moving Pictures Experts Group* (MPEG), conhecidos como H.264, ou, Codificação de Vídeo Avançada (AVC), e, H.265, ou, Codificação de Vídeo de Alta Eficiência (HEVC). O padrão mais recente, H.265/HEVC, visa reduzir a taxa de bits pela metade em comparação com seu antecessor, para uma mesma qualidade de imagem. O desenvolvimento deste recente padrão em codificação de vídeo objetiva atender futuros desafios impostos sobre as redes de comunicação e dispositivos diversos reprodutores de vídeo, baseado numa demanda sempre crescente por qualidade e de reduzida taxa de bits. Tal padronização foi projetada de forma a estar preparada para lidar com resoluções Ultra HD, resoluções conhecidas como 4k e 8k. Os resultados demonstram a eficiência do padrão H.265/HEVC, principalmente para taxa de bits maiores que 750 kbps. Contudo, o custo computacional ainda é elevado, em pelo menos 3 vezes mais em relação ao padrão H.264/AVC.

**Palavras-chave:** Codificação de vídeo. H.264/AVC. H.265/HEVC.

## ABSTRACT

The current work performs a objective comparative analysis in video coding. The subjects of study are the standards created by International Telecommunication Union – Telecommunication Sector (ITU-T) and ISO/IEC Moving Pictures Experts Group (MPEG), known as H.264, or, Advanced Video Coding (AVC), and, H.265, or, High Efficient Video Coding (HEVC). The most recent standard, the H.265/HEVC, aims to halve the bitrate compared to its predecessor, for the same image quality. The development of this video coding standard focus to attend future needs placed on the transmission ways and diverse video-reproducer devices, based on a demand ever increasing for quality and diminished bitrate. Such standardization was designed in a way to be prepared to handle Ultra HD resolutions, known as 4k and 8k. The results show the efficiency of H.265/HEVC standard, with emphasis for greater bitrates than 750 kbps. However, the computational cost is still high, with 3 times higher, at least, relative to H.264/AVC.

**Keywords:** Video Coding. H.264/AVC. H.265/HEVC.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Formatos de Sub-amostragem.....	14
Figura 2 – Codificador híbrido de vídeo.....	19
Figura 3 – Etapas do processo de codificação/decoficação.....	20
Figura 4 - Subdivisão de uma imagem em fatias (quando não usando FMO).....	24
Figura 5 - Esquerda: predição Intra 4x4 é aplicada para amostras a-p utilizando amostras A-M. Direita: Oito "direções de predição" selecionáveis para modo Intra 4x4.....	27
Figura 6 - Segmentações de macroblocos para MC. Topo: segmentação de macroblocos, Base: segmentação de partições (sub-macroblocos) 8x8.....	28
Figura 7 – 4 imagens previamente decodificadas como referência.....	29
Figura 8 – Desempenho do filtro de deblocação para imagens altamente comprimidas. Esq.: Sem filtragem; Dir.: Com filtragem.....	32
Figura 9 – Ilustração dos perfis do padrão H.264/AVC.....	33
Figura 10 – Níveis do padrão H.264/AVC.....	34
Figura 11 – Diferentes tipos de estrutura de predição hierárquica utilizada na escalabilidade temporal. Os números abaixo de cada imagem especificam a ordem de codificação e, os símbolos T(k) especificam as camadas temporais com k como identificadores de cada uma.....	37
Figura 12 – Estrutura multicamada com camada intercalada adicional utilizada na escalabilidade espacial.....	37
Figura 13 - Um exemplo da estrutura de um codificador SVC.....	39
Figura 14 – Subdivisão de um CTB em CBs (e <i>Transform Blocks</i> - TBs). Linhas contínuas indicam as arestas dos CBs; linhas pontilhadas indicam arestas de TBs (Lado esq. Particionamento do CTB; Lado dir. <i>quadtree</i> correspondente).....	42
Figura 15 - Modos para divisão de um CB em PBs, sujeita a certas limitações de tamanho. Para CBs preditos por predição <i>Intra</i> , apenas M x M e M/2 x M/2 são suportados.....	43
Figura 16 – Modos e orientações direcionais para predição <i>Intra</i> .....	45
Figura 17 – Níveis do padrão H.265/HEVC.....	47
Figura 18 – Subdivisão de uma imagem em (a) fatias, (b) <i>tiles</i> , (c) Ilustração do processamento de frente de onda.....	50

Figura 19 – Sequência Foreman em formato YUV.....	60
Figura 20 – Sequência Miss-américa em formato YUV.....	61
Figura 21 – Sequência Salesman em formato YUV.....	61
Figura 22 – Testes e taxas de bits alvo utilizadas.....	65
Figura 23 – Gráfico PSNR-Y (db) x Taxa de Bits (kbps) para a sequência Foreman	68
Figura 24 – Quadros decodificados da sequência Foreman: (a) H.264/AVC 32 kbps (b) H.264/AVC 1024 kbps (c) H.265/HEVC 32 kbps (d) H.265/HEVC 1024 kbps.....	68
Figura 25 – Quadro original para a sequência Foreman.....	69
Figura 26 – Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Foreman.....	69
Figura 27 – Gráfico PSNR-Y x Taxa de Bits para a sequência Miss-américa.....	70
Figura 28 - Quadros decodificados da sequência Miss-américa: (a) H.264/AVC 32 kbps (b) H.264/AVC 1024 kbps (c) H.265/HEVC 32 kbps (d) H.265/HEVC 1024 kbps. .....	70
Figura 29 – Quadro original para a sequência Miss-américa.....	71
Figura 30 - Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Miss-américa.....	71
Figura 31 – Gráfico PSNR-Y x Taxa de Bits para a sequência Salesman.....	72
Figura 32 - Quadros decodificados da sequência Salesman: (a) H.264/AVC 32 kbps (b) H.264/AVC 1024 kbps (c) H.265/HEVC 32 kbps (d) H.265/HEVC 1024 kbps.....	73
Figura 33 – Quadro original para a sequência Salesman.....	73
Figura 34 - Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Salesman.....	74
Figura 35 – Análise dos resultados de desempenho para a sequência Foreman.....	76
Figura 36 – Análise dos resultados de desempenho para a sequência Miss-américa. .....	76
Figura 37 – Análise dos resultados de desempenho para a sequência Salesman....	77

## LISTA DE TABELAS

Tabela 1 – Lista de sequências utilizadas nos testes e suas características.....	60
Tabela 2 – Parâmetros de execução fornecidos ao codificador do software <i>JM Reference Model</i> .....	64
Tabela 3 – Parâmetros de execução fornecidos ao codificador do software <i>HM Reference Model</i> .....	64

## LISTA DE ABREVIATURAS E SIGLAS

AMVP	<i>Advanced Motion Vector Prediction</i>
ASO	<i>Arbitrary Slice Ordering</i>
AVC	<i>Advanced Video Coding</i>
CABAC	<i>Context-Adaptive Binary Arithmetic Coding</i>
CAVLC	<i>Context-Adaptive Variable-Length Coding</i>
CBs	<i>Coding Blocks</i>
CTBs	<i>Coding Tree Blocks</i>
CTU	<i>Coding Tree Unit</i>
CU	<i>Coding Unit</i>
DCT	<i>Discrete Cosine Transform</i>
DP	<i>Direct Prediction</i>
DPB	<i>Decoded Picture Buffer</i>
FIR	<i>Finite Impulse Response</i>
FMO	<i>Flexible Macroblock Ordering</i>
FRExt	<i>Fidelity Range Extensions</i>
HD	<i>High Definition</i>
HEVC	<i>High Efficient Video Coding</i>
HP	<i>High Profile</i>
ITU-T	<i>International Telecommunication Union – Telecommunication sector</i>
JCT	<i>Joint Collaborative Team</i>
KLT	<i>Karhunen-Loève Transform</i>
MC	<i>Motion Compensation</i>
MCP	<i>Motion Compensation Prediction</i>
ME	<i>Motion Estimation</i>
MP	<i>Main Profile</i>
MPEG	<i>Moving Pictures Experts Group</i>
MSE	<i>Mean-Squared Error</i>
MVs	<i>Motion Vectors</i>
MTU	<i>Maximum Transfer Unit</i>
NAL	<i>Network Abstraction Layer</i>
PBs	<i>Prediction Blocks</i>
PU	<i>Prediction Units</i>

QP	<i>Quantization Parameter</i>
SEI	<i>Supplemental Enhancement Information</i>
SNR	<i>Signal-to-Noise Ratio</i>
TBs	<i>Transform Blocks</i>
TUs	<i>Transform Units</i>
URQ	<i>Uniform Reconstruction Quantization</i>
UNESC	Universidade do Extremo Sul Catarinense
VCL	<i>Video Coding Layer</i>
VCEG	<i>Video Coding Experts Group</i>
VLC	<i>Variable Length Code</i>
VoD	<i>Video on Demand</i>
VQA	<i>Subjective Video Quality Assesment</i>
VUI	<i>Video Usability Information</i>
WPP	<i>Wavefront Parallel Processing</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>8</b>
1.1 OBJETIVO GERAL.....	10
1.2 OBJETIVOS ESPECÍFICOS.....	10
1.3 JUSTIFICATIVA.....	11
1.4 ESTRUTURA DO TRABALHO.....	12
<b>2 CODIFICAÇÃO DE VÍDEO</b> .....	<b>13</b>
2.1 ESPAÇO E SUB-AMOSTRAGEM DE CORES.....	13
2.2 CARACTERÍSTICAS DETERMINANTES DAS CODEC(S).....	15
2.3 TÉCNICAS PARA COMPRESSÃO DIGITAL.....	15
2.4 EXPLORANDO A REDUNDÂNCIA DE INFORMAÇÕES EM VÍDEOS.....	16
2.5 CODEC(S) MODERNAS DE VÍDEO.....	17
2.6 ETAPAS DA CODIFICAÇÃO DE VÍDEO.....	18
<b>3 PADRÃO H.264/AVC</b> .....	<b>21</b>
3.1 CAMADA DE ABSTRAÇÃO DE REDE.....	21
<b>3.1.1 Unidades NAL</b> .....	<b>22</b>
<b>3.1.2 Parâmetros de configuração</b> .....	<b>22</b>
<b>3.1.3 Unidades de Acesso</b> .....	<b>23</b>
3.2 CAMADA DE VÍDEO CODIFICADO.....	23
<b>3.2.1 Macroblocos, Fatias, e Grupos de Fatias</b> .....	<b>23</b>
<b>3.2.2 Tipos de Fatias</b> .....	<b>26</b>
<b>3.2.3 Predição Intra de fatias</b> .....	<b>26</b>
<b>3.2.4 Predição Inter de fatias</b> .....	<b>28</b>
3.2.4.1 Fatias P.....	28
3.2.4.2 Predição Inter em fatias B.....	29
3.2.4.3 Predição ponderada em fatias P e B.....	30
<b>3.2.5 Transformação, Escalamento e Quantização</b> .....	<b>30</b>
<b>3.2.6 Codificação por entropia</b> .....	<b>31</b>
<b>3.2.7 Filtro de deblocagem</b> .....	<b>31</b>
<b>3.2.8 Perfis e níveis</b> .....	<b>32</b>
3.2.8.1 Perfis: Baseline, Main e Extended (Versão 1).....	34
3.2.8.2 Perfis <i>High</i> da emenda FRExt (Versão 3).....	35
3.3 ESCALABILIDADE.....	36
<b>3.3.1 Vídeos escaláveis</b> .....	<b>36</b>

<b>3.3.2 A extensão SVC</b> .....	<b>38</b>
<b>4 PADRÃO H.265 HEVC</b> .....	<b>40</b>
4.1 DESIGN E RECURSOS.....	40
<b>4.1.1 Camada de Codificação de Vídeo (VCL)</b> .....	<b>40</b>
4.1.1.1 Coding Tree Units, Coding Tree Blocks, Coding Units, e Coding Blocks.....	41
4.1.1.2 Prediction Units e Prediction Blocks.....	42
4.1.1.3 Transform Units e Transform Blocks.....	43
4.1.1.4 Representação de Vetores de Movimento.....	43
4.1.1.5 Compensação de Movimento.....	44
4.1.1.6 Predição <i>Intra</i> .....	44
4.1.1.7 Controle de quantização.....	45
4.1.1.8 Codificação por entropia.....	45
4.1.1.9 Filtro de deblocagem <i>in-loop</i> (aninhado em laço).....	45
4.1.1.10 Perfis e níveis.....	46
<b>4.1.2 Camada de Abstração de Rede (NAL)</b> .....	<b>47</b>
4.1.2.1 Estrutura dos parâmetros de configuração.....	47
4.1.2.2 Estrutura da sintaxe da unidade NAL.....	48
4.1.2.3 Fatias.....	48
4.1.2.4 SEI e VUI.....	48
<b>4.1.3 Elementos de arquitetura de processamento paralelo</b> .....	<b>48</b>
4.1.3.1 Tiles.....	48
4.1.3.2 Processamento de frente-de-onda.....	49
4.1.3.3 Segmentos de fatia dependente.....	49
<b>5 TRABALHOS CORRELATOS</b> .....	<b>51</b>
5.1 ANÁLISE COMPARATIVA ENTRE OS CODIFICADORES DE VÍDEO H.264/AVC E H.265/HEVC.....	51
5.2 QUANTITATIVE PERFORMANCE EVALUATION OF THE EMERGING HEVC/H.265 VIDEO CODEC.....	52
5.3 PERFORMANCE COMPARISON OF H.265/MPEG-HEVC, VP9, AND H.264/MPEG-AVC ENCODERS.....	52
5.4 MPEG VIDEO DEPLOYMENT IN INTERACTIVE MULTIMEDIA SYSTEMS: HEVC VS. AVC CODEC PERFORMANCE STUDY.....	53
5.5 AVALIAÇÃO DE DESEMPENHO DA ESCALABILIDADE ESPACIAL NA RECOMENDAÇÃO H.264/SVC PARA CODIFICAÇÃO DE VÍDEO.....	54

5.6	SEGMENTAÇÃO QUADTREE PARA SELEÇÃO DE BLOCOS NA	54
	CODIFICAÇÃO DE VÍDEO H.264/AVC	54
5.7	DISTRIBUIÇÃO DE VÍDEO ESCALÁVEL BASEADA NA RECOMENDAÇÃO	
	H.264/SVC EM UM SERVIÇO DE WEBTV	54
<b>6</b>	<b>ANÁLISE COMPARATIVA ENTRE OS PADRÕES ITU-T H.265/HEVC E</b>	
	<b>H.264/AVC EM CODIFICAÇÃO DE VÍDEO</b>	<b>56</b>
6.1	METODOLOGIAS EM ANÁLISES DE CODIFICAÇÃO DE VÍDEO	57
<b>6.1.1</b>	<b>Abordagem Objetiva</b>	<b>57</b>
6.1.1.1	Taxa de bits	57
6.1.1.2	Complexidade Computacional	58
6.1.1.3	Razão Sinal-Ruído de Pico (PSNR)	58
<b>6.1.2</b>	<b>Abordagem Subjetiva</b>	<b>59</b>
<b>6.1.3</b>	<b>Metodologia empregada no presente trabalho</b>	<b>59</b>
6.1.3.1	Seleção e características das sequências de vídeo utilizadas para teste	60
6.1.3.2	Coleta de dados	62
6.2	CENÁRIO DA ANÁLISE COMPARATIVA EFETUADA	62
<b>6.2.1</b>	<b>Implementações utilizadas</b>	<b>62</b>
<b>6.2.2</b>	<b>Ambiente de execução e configurações iniciais de software</b>	<b>63</b>
<b>6.2.3</b>	<b>Parâmetros de execução e arquivos de configuração</b>	<b>63</b>
6.3	DESAFIOS ENCONTRADOS	66
6.4	RESULTADOS	67
6.5	ANÁLISE DOS RESULTADOS	75
6.6	DISCUSSÃO DA EFETIVIDADE DAS MEDIÇÕES EFETUADAS E MÉTRICAS	78
<b>7</b>	<b>CONCLUSÃO</b>	<b>79</b>
	<b>REFERÊNCIAS</b>	<b>80</b>
	<b>APÊNDICE A - ARTIGO</b>	<b>86</b>

## 1 INTRODUÇÃO

Na última década e meia, tecnologias de compressão de vídeo digital se tornaram uma parte fundamental do modo como se cria, comunica e se consome informação visual. Os avanços no desempenho das codificações de vídeo devem-se a implementação de diferenciadas abordagens e estratégias que buscam a otimização computacional para vídeos digitais (SULLIVAN; WIEGAND, 2005, tradução nossa).

Utilizam-se variadas técnicas de compressão de sinal para efetuar a diminuição do tamanho dos arquivos de vídeo a serem transmitidos ou armazenados. Codificadores modernos procuram empregar estas várias técnicas cumulativamente, caracterizando-se como tecnologias híbridas de compressão de vídeos digitais. É importante ressaltar a relevância do uso de codificações de vídeo para efetuar a compressão deste tipo de arquivo, que caracteriza-se por conter considerável quantidade de informação. Arquivos de vídeo em formato original/planar, ou seja, sem compressão, possuem alto grau de redundância no que se refere à representação de cada pixel constituinte de cada quadro da sequência do vídeo. Sendo assim, a adoção da compressão de tais informações é a forma mais eficaz de tratar arquivos de vídeo, em termos de armazenamento, transmissão e reprodução. Do contrário, seria inviável, por exemplo, transmitir vídeos pela Internet ou outro meio de transmissão, visto que a velocidade de envio dos dados seria desproporcional ao tamanho dos arquivos. Até mesmo serviços de vídeo que não requerem baixa latência, seriam impraticáveis.

Atualmente, é comum o uso do termo 'alta definição' (HD), referindo-se a serviços e dispositivos que suportam resoluções iguais ou superiores a 720 pixels. Um crescente número de serviços de *streaming* e de vídeo sob demanda (VoD) tem disponibilizado amplamente conteúdo de alta definição. Não obstante, o mercado de dispositivos móveis tem se difundido globalmente a passos largos, oferecendo dispositivos para a população em geral, capazes de reproduzir e criar conteúdo visual em qualidade equivalente.

Além disso, a busca por qualidades superiores ao reproduzir vídeos digitais, é uma tendência que futuramente criará a necessidade de suplantando as resoluções atuais por outras ainda maiores, como as resoluções 4k e 8k. Como efeito decorrente desta demanda, impõe-se um custo computacional maior para

sistemas de compressão e reprodução de vídeos digitais, assim como desafios para os meios de transmissão (SULLIVAN et al, 2012, tradução nossa).

A *International Telecommunication Union – Telecommunication sector* (ITU-T), têm desenvolvido padrões de codificação ao longo dos anos que têm proporcionado a regulamentação de codificadores e decodificadores de vídeo, objetivando as demandas em termos de latência e qualidade. O aumento da capacidade de processamento ao longo do tempo tem permitido atacar o problema por diferentes ângulos. É notável um desenvolvimento em estrutural e de robustez no processamento de vídeos a cada novo padrão criado, com ênfase para os padrões mais recentes, H.264/AVC e H.265/HEVC.

O padrão de Codificação de Vídeo Avançada (AVC), ou simplesmente H.264 – publicado no ano de 2003 -, está atualmente em vigor, presente em diversos tipos de serviços de reprodução de vídeo digital. O processo de codificação descrito por este padrão faz uso de modelos matemáticos para predição e otimização da informação visual presente num vídeo digital. A sua arquitetura está dividida em duas camadas, uma de rede – Camada de Abstração de Rede, responsável pela formatação e envio dos dados pelos meios de transmissão - e, uma camada responsável por lidar com a codificação de vídeo propriamente dita – Camada de Codificação de Vídeo. Ambas possuem vários níveis de abstração e interdependência de dados que visam robustez e a resiliência a erros. No entanto, sendo previsível que o padrão H.264/AVC também terá um prazo de efetividade, a ITU-T *Video Coding Experts Group* (VCEG) lançou-se no desenvolvimento de um novo padrão, junto ao grupo ISO/IEC *Moving Pictures Experts Group* (MPEG), criando o *Joint Collaborative Team* (JCT) (SULLIVAN; WIEGAND, 2005, tradução nossa).

O padrão de Codificação de Vídeo de Alta Eficiência (HEVC), ou simplesmente H.265 – publicado no ano de 2013 -, toma como base o seu antecessor, mantendo diversos elementos estruturais e implementando novas abordagens para o processo de codificação. É sabido que esta última implementação da ITU-T/MPEG, reduz pela metade o fluxo de dados *versus* um aumento de três vezes mais em custo computacional (KOUMARAS et al, 2012, tradução nossa).

Enquadra-se dentro do escopo deste trabalho de conclusão de curso, efetuar uma análise comparativa, do tipo objetiva/quantitativa, entre os padrões de codificação citados. Tal empreitada se deu por meio de testes de codificação

realizados com sequências de vídeo em formato planar. Durante os testes, foram armazenados os dados e informações resultantes, de forma a prover uma comparação baseada em métricas quantitativas largamente empregadas em estudos em codificação de vídeo. A partir destas, foi possível avaliar o desempenho e a complexidade computacional dos arquivos codificados. Foram selecionados e utilizados softwares que implementam os padrões em análise.

## 1.1 OBJETIVO GERAL

Avaliar o desempenho do Padrão H.265/HEVC comparativo ao padrão H.264/AVC para as características de Taxa de Bits, Razão Sinal-Ruído de Pico (PSNR) e Complexidade Computacional.

## 1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos são:

- a) descrever as características do padrão H.264/AVC ;
- b) descrever as características do padrão H.265/HEVC;
- c) descrever as diferenças no processo de codificação de vídeo para os padrões H.264/AVC e H.265/HEVC;
- d) descrever as *codecs* baseadas no padrão H.264/AVC e no padrão H.265/HEVC para o desenvolvimento do estudo comparativo;
- e) aplicar as *codecs* em estudo de caso para avaliação de desempenho e complexidade computacional, a partir de métricas quantitativas de taxa de bits, Razão Sinal-Ruído de Pico (PSNR) e Complexidade Computacional.

### 1.3 JUSTIFICATIVA

A popularização de vídeos de alta definição e o surgimento de formatos maiores que o HD, demandam tecnologias de codificação de vídeo mais eficientes que as capacidades do H.264/AVC. Isto se torna ainda mais perceptível quando resoluções de maior tamanho são acompanhadas por exibições estéreo ou captura *multiview*. Não obstante, o tráfego causado por aplicações de vídeo para dispositivos móveis e *tablets*, assim como os requisitos de transmissão para serviços de vídeo sob demanda, estão impondo desafios severos nas redes atuais. Um desejo crescente por maior qualidade e resolução está, também, crescendo em aplicações de dispositivos móveis (SULLIVAN et al, 2012, tradução nossa).

Previsões sobre o tráfego de rede, como as feitas pela *Cisco Visual Networking Index*, prevêm forte crescimento nas taxas de tráfego de vídeo. Previsões de crescimento anual esperam taxas de 30% ou mais para serviços de vídeo baseados em IP por cabo e 90% para a InternetTV em redes de dispositivos móveis. Devido a estas altas de crescimento, a transmissão de vídeo contabilizará como uma grande porção do tráfego nas redes. A Cisco, Inc., previu que a transmissão de vídeo contribuirá aproximadamente a dois terços do tráfego de rede de dispositivos móveis em 2014 (SEELING; REISSLEIN, 2014, tradução nossa).

O H.265/HEVC foi projetado para lidar essencialmente com todas as aplicações do H.264/AVC e, focar em especial, em dois problemas chave: aumento da resolução de vídeo e aumento de uso de arquiteturas de processamento paralelo. Sua sintaxe é genérica e foi desenvolvida de forma a se adaptar inclusive para aplicações específicas não mencionadas acima. O padrão foi desenvolvido para alcançar múltiplos objetivos, incluindo eficiência de codificação, facilidade na integração de sistemas de transporte e resiliência à perda de dados, assim como, implementabilidade através de arquiteturas paralelas de processamento (SULLIVAN et al, 2012, tradução nossa).

Desta forma, o presente trabalho teve a intenção de prover dados científicos úteis na avaliação do estado de arte das tecnologias em codificação de vídeo.

## 1.4 ESTRUTURA DO TRABALHO

Procurou-se neste trabalho, primeiramente, explicar sobre assuntos pertinentes à codificação de vídeo (Capítulo 2). Da mesma forma, a estrutura e funcionamento dos padrões aos quais esta pesquisa se atém, foram elucidados de forma a obter-se uma noção sobre dois padrões referência em codificação de vídeos da atualidade (Capítulos 3 e 4). Logo em seguida, tomou-se lugar uma listagem de trabalhos correlatos (Capítulo 5).

Num segundo momento, uma comparação analítica baseada em métricas objetivas-quantitativas, consensualmente empregadas em estudos em codificação de vídeo, foi efetuada fazendo-se uso das versões mais atuais dos modelos de teste oficiais (Capítulo 6).

Finalmente, conclusões e considerações finais encontram-se ao final deste trabalho (Capítulo 7).

## 2 CODIFICAÇÃO DE VÍDEO

Para a codificação de vídeo, define-se os seguintes objetivos: transmitir vídeos usando a menor taxa de bits possível, mantendo-se a fidelidade de reprodução em comparativo ao vídeo original (SULLIVAN; WIEGAND, 2005, tradução nossa).

Um sistema de codificação que cumpre com tais expectativas, possui: eficiência de codificação e taxas de distorção aceitáveis em relação ao vídeo original. Tal sistema pode ser chamado de *codec*, ou seja, um sistema que engloba um codificador e um decodificador (SULLIVAN; WIEGAND, 2005, tradução nossa).

A título de esclarecimento, ressalta-se que ao longo deste trabalho foram apenas abordadas *codecs* de vídeo – no caso, as referentes aos padrões H.264/AVC e H.265/HEVC – que se enquadram no tipo 'com-perda', pois removem certas partes do vídeo original, gerando arquivos decodificados que não são idênticos ao seu original. Tal estratégia de codificação visa obter altas taxas de compressão, removendo informações redundantes que não necessariamente serão detectadas pela visão humana. No entanto, as seguintes seções 2.2 e 2.3 abordarão conceitos comuns a codificadores de sinais em geral<sup>1</sup>.

### 2.1 ESPAÇO E SUB-AMOSTRAGEM DE CORES

As cores são originalmente representadas pelo sistema de cores RGB, o qual é composto por três cores: vermelha, verde e azul. Devido ao seus componentes de cor possuir alto grau de correlação, o processamento independente de tais se torna dificultado (DIAS; ALCAIM, 2006).

Sendo assim, o sistema de cores RBG é convertido em outro sistema, que possui certas diferenças. O espaço de cores YCbCr, é também utilizado para representação de imagens digitais, sendo constituído pelos seguintes componentes: luminância (Y), que determina a intensidade luminosa - ou de luminescência - , crominância azul (Cb) e crominância vermelha (Cr).

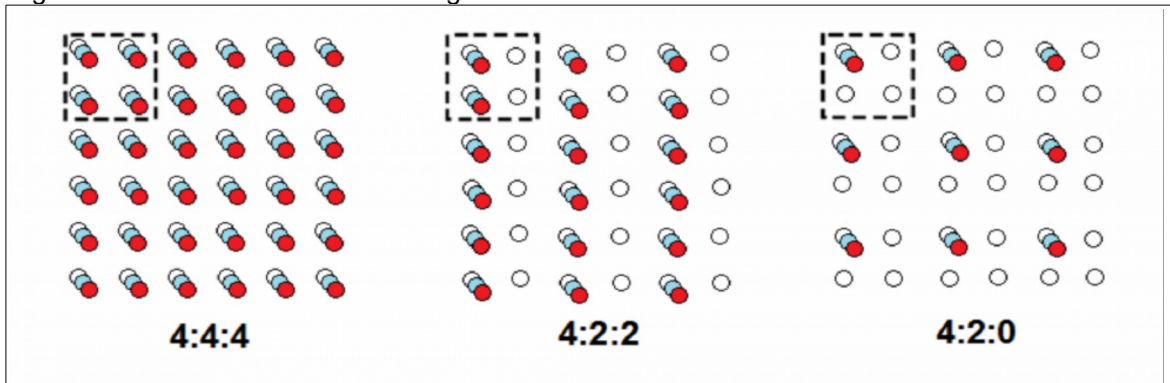
---

<sup>1</sup>Note-se que o longo deste capítulo e do próximo foi feito uso extensivo da referência Sullivan e Wiegand (2005). Deste modo, procura-se evitar a repetitividade na indicação do embasamento utilizado, sendo referenciado apenas materiais complementares para o desenvolvimento deste capítulo.

A vantagem de separar o componente de luminosidade do das cores, é permitir representar imagens digitais de acordo com as capacidades visuais humanas. É sabido que a visão humana é muito mais sensível à luminosidade do que a crominância. Dessa forma, codificadores tomam proveito desta característica, reduzindo a quantidade de informação dedicada à representação de cores e dando mais ênfase à luminescência.

Tal sub-amostragem da cor pode reduzir o tamanho de uma imagem em 25-50%. Existem diferentes formatos de sub-amostragem, os quais criam diferentes proporções entre luminosidade e as crominâncias. Abaixo, é possível observar na figura 1 alguns exemplos (DIAS; ALCAIM, 2006).

Figura 1 – Formatos de Sub-amostragem.



Fonte: Silva (2014, p.9).

Na figura acima pode-se ver a representação do componente Y e os componentes Cb e Cr, como círculos branco, azul e vermelho, respectivamente. A série de 3 números separados por ':' - 4:4:4 , 4:2:2 e 4:2:0 – são as proporções mencionadas anteriormente. A terceira sub-amostragem de cores, da esquerda para a direita, é largamente adotada e foi o formato adotado para este trabalho. No entanto, é intuitivo dizer que os dois formatos anteriores possuem uma acurácia maior na representação das cores de uma imagem, pelo simples fato de dedicarem mais amostras para cada bloco amostrado.

## 2.2 CARACTERÍSTICAS DETERMINANTES DAS CODEC(S)

*Codecs* são essencialmente caracterizadas em função de:

- a) taxa de transferência do canal: uma característica influenciada pela taxa de bits e a sobrecarga suportada pelo sistema de transmissão causada por protocolo(s) e correções de erro de codificação;
- b) distorção do arquivo de saída: distorção é principalmente induzida pela *codec* e pelos erros do canal introduzidos no 'caminho' para o decodificador.

Contudo, na prática, a transmissão de vídeo deve-se considerar, também, os seguintes problemas:

- c) *delay*: que se refere a latência de início e *end-to-end delay*. As características de *delay* são influenciadas por muitos parâmetros, incluindo *buffering*, *delays* estruturais da *codec*, e a velocidade que os dados são transmitidos através do sistema de transmissão;
- d) complexidade: em termos computacionais, capacidade de memória e requerimentos de acesso de memória.

## 2.3 TÉCNICAS PARA COMPRESSÃO DIGITAL

Existem diferentes técnicas de compressão de um sinal digital, os quais são também utilizados em codificação de arquivos de vídeo digital.

A predição, um processo pelo qual um conjunto de valores preditos é criado, é normalmente baseado numa indicação enviada pelo codificador sobre como proceder em tal situação. Sendo assim, tal processo é baseado na análise de amostras de entrada e os tipos de predição que podem ser selecionados na arquitetura do sistema. A finalidade é prever os valores de entrada futuros, de forma que os valores que precisam ser atualizados se tornam apenas as diferenças em relação aos valores preditos. Estas diferenças são chamadas de valores residuais.

A transformação, um processo (também chamado de decomposição de sub-banda) que está diretamente relacionado à predição, consiste de formar novos conjuntos de amostras a partir de uma combinação de entrada, frequentemente usando combinação linear. Em termos gerais, uma transformação pode prevenir a necessidade de representar valores similares repetidamente e capturar a “essência”

do sinal de entrada, utilizando uma análise de frequência. Um benefício típico da transformação, é a redução na correlação estatística das amostras de entrada, de forma que os aspectos mais relevantes do conjunto das amostras de entrada são usualmente concentrados em um número menor de variáveis. Dois bons exemplos de transformação são efetuados pelas Transformada Karhunen-Loève (KLT) e Transformada Discreta do Coseno (DCT).

A quantização, um processo pelo qual a precisão usada para a representação de um valor de amostragem ou de um grupo de valores de amostragem, é reduzida, com intenção de diminuir a quantidade de dados necessários para codificação. Tal processo é análogo ao conceito conhecido de arredondamento de dígitos menos significativos ao escrever algum valor estatístico. Entre as técnicas aqui listadas para compressão, a quantização é geralmente a única que caracteriza-se por ser não-reversível – visto que envolve uma forma de mapear 'muitos-para-alguns', caracterizando-se por certa perda de fidelidade. O desafio é minimizar essa perda de fidelidade usando como parâmetro algum método de medição de distorção.

A codificação por entropia, um processo pelo qual símbolos discretos de entrada são representados de uma forma que se torna vantajoso em relação as probabilidades relativas a vários valores possíveis para cada símbolo fonte. Um tipo conhecido de codificação é o Código de Comprimento Variável (VLC), o qual envolve estabelecer uma tabela de código de árvore estruturada que usa strings binárias curtas para representar valores menos frequentes. O melhor método para implementar VLCs é o conhecido método do código de Huffman, que produz um VLC otimizado. Outro método, de certa forma menos conhecido, potencialmente mais otimizado e mais facilmente projetado para adaptar-se a símbolos estatísticos variantes do que a codificação VLC, é a técnica chamada de codificação Aritmética.

## 2.4 EXPLORANDO A REDUNDÂNCIA DE INFORMAÇÕES EM VÍDEOS

Os codificadores de vídeo procuram reduzir a taxa de transmissão de bits para que a transmissão seja possível nos meios de comunicação. Caso contrário, a quantidade total de bits seria muito maior e a transmissão se tornaria impraticável. Sendo assim, é feita a remoção de informações redundantes dos sinais dos vídeos. Por definição, existem três tipos de redundâncias: espacial, temporal e psico-visual (DIAS; ALCAIM, 2006).

Em um determinado quadro durante a reprodução de um vídeo, é possível determinar a crominância e a luminescência de pixels vizinhos por aproximação, uma vez que possuem semelhança nestas características (redundância espacial). Dessa forma, as *codecs* buscam reescrever informações de quadros consecutivos baseados nessas informações, economizando bits para representar cada pixel. Essa técnica resulta em redução da qualidade da imagem, imperceptível, no entanto, à visão humana. Esta redundância é explorada pela predição *intra-picture* e sinais codificados neste modo são ditos como *Intra-codificados* (DIAS; ALCAIM, 2006).

Não obstante, muito da cena apresentada é praticamente apenas repetida quadro após quadro sem nenhuma mudança significativa, de forma que o vídeo pode ser representado mais eficientemente apenas enviando as mudanças na cena ao longo do tempo, em vez de codificar todas as regiões repetidamente.

A “movimentação” de pixels ao longo da execução de quadros de um vídeo, pode, também, ser predita. É comum que durante um vídeo, certos pixels tendem a ficar imóveis, sem alteração. Dessa forma, a *codec* pode reescrever informação baseada em quadros passados ou até mesmo futuros. Esta redundância temporal é explorada pela predição *Inter-prediction* e sinais codificados neste modo são ditos como Inter-codificados (DIAS; ALCAIM, 2006).

Para a redundância psico-visual temos o sistema de visão humano, o qual apresenta limitações para distinguir detalhes e acompanhar imagens em movimentos rápidos. As *codecs* aproveitam-se disto reduzindo a qualidade do vídeo de forma que seja imperceptível ao olho humano (DIAS; ALCAIM, 2006).

## 2.5 CODEC(S) MODERNAS DE VÍDEO

Atualmente, *codecs* podem ser tidas como híbridas por implementarem duas técnicas de redução de redundância, a de predição e de transformação. Em *codecs* modernas, regiões de cada quadro podem ser preditas utilizando a predição *Inter-picture* e uma transformada espacial baseada em frequência e *Intra-codificada*, aplicada nas regiões de refinamento.

O conceito de Predição de Compensação de Movimento (MCP), têm sido a principal causa nos avanços de codificação de vídeo. A maior parte das mudanças

no conteúdo de vídeos é normalmente devida a movimentação de objetos na cena apresentada relativamente ao plano da imagem. Uma pequena quantidade de movimento pode resultar em grandes diferenças nos valores das amostras de uma imagem, especialmente perto da borda/limite de objetos. Frequentemente, prever uma região de uma imagem a partir de uma imagem antecessora, movimentada em sua localização espacial, pode significativamente reduzir a necessidade por um refinamento das mudanças na imagem. O uso de Vetores de Movimentação (MVs) para formar uma previsão baseada no deslocamento espacial, é conhecido como Compensação de Movimento (MC).

É sabido que melhorias subsequentes de técnicas MCP foram a maior razão para melhoria na eficiência de codificação, alcançados por padrões modernos quando comparados com gerações anteriores. O custo computacional pelo uso da técnica de MCP, de forma cada vez mais sofisticada, é o fator principal para o aumento nos requisitos de complexidade.

## 2.6 ETAPAS DA CODIFICAÇÃO DE VÍDEO

Em suma, geralmente, um algoritmo de codificação de vídeo funciona da seguinte maneira: cada imagem é dividida em blocos. A primeira imagem da sequência de um vídeo é *Intra*-codificada. Para todas as imagens restantes desta sequência de quadros, modos de codificação *Inter-picture* são usados para descobrir a maior parte dos blocos. É neste processo em que ocorre a Estimativa de Movimento (ME), que consiste em escolher dados em movimento utilizando a imagem selecionada como referência e, MVs para serem aplicados para cada amostra de cada bloco.

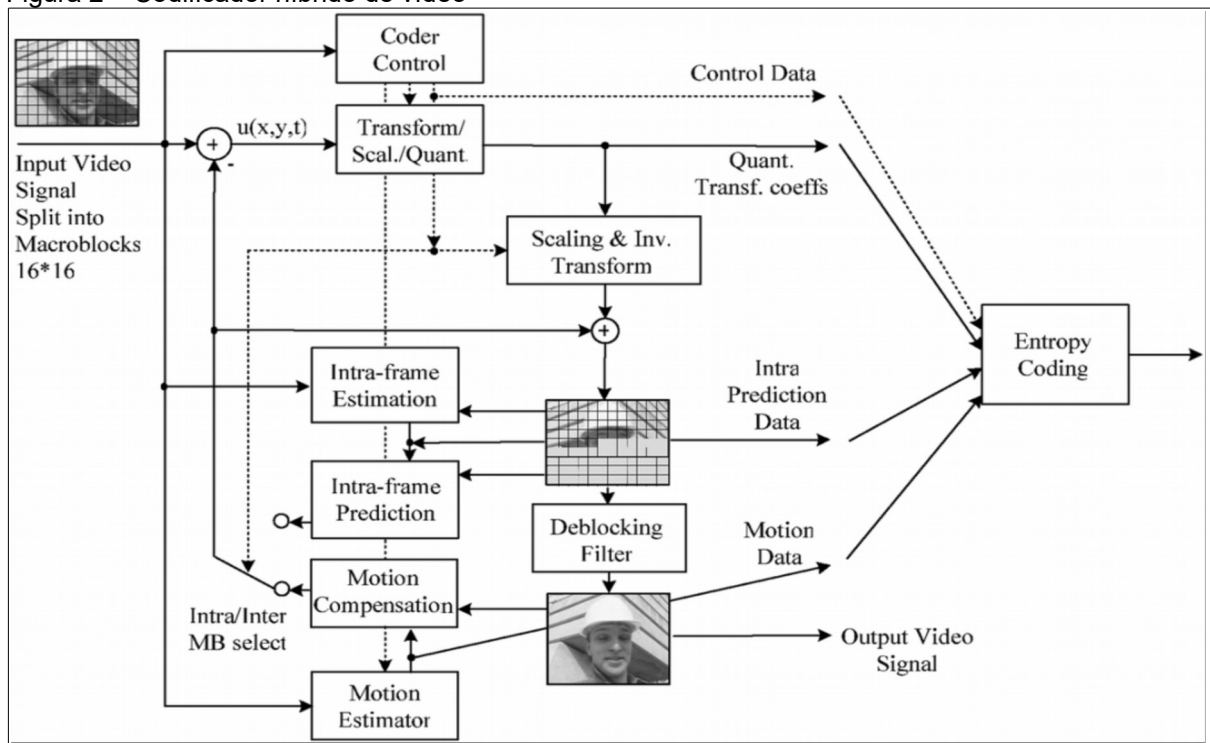
São transmitidos como informação complementar, a descrição do movimento e dados úteis para determinar o modo em que se codificará - que são usados por ambos codificador e decodificador - para gerar a previsão de forma fiel aos quadros originais, fazendo uso da MC. O residual das previsões *Intra* e *Inter*, o qual é a diferença entre o bloco original e a sua previsão, é transformado por uma transformada de frequência. Os coeficientes da transformada são então escalados, quantizados, codificados por entropia, e transmitidos juntos com a informação extra de previsão. O codificador duplica o processo de decodificação, de forma que, ambos vão gerar previsões idênticas para dados subsequentes. Portanto, os coeficientes transformados e quantizados são construídos proporcionalmente

invertidos e são então transformados inversamente para duplicar a predição residual decodificada.

O resíduo é então adicionado à predição, e o resultado de tal adição poderá ser alimentado num filtro de suavização de blocos para atenuar as descontinuidades criadas por suas arestas, induzidas no processamento de blocos. A imagem final (que é também gerada pelo decodificador) é então armazenada para a predição de imagens codificadas subsequentes.

Finalmente, o design e a operação de um codificador envolve a otimização de muitas decisões para alcançar a melhor relação possível entre taxa e distorção, dadas as restrições para *delay* e complexidade.

Figura 2 – Codificador híbrido de vídeo

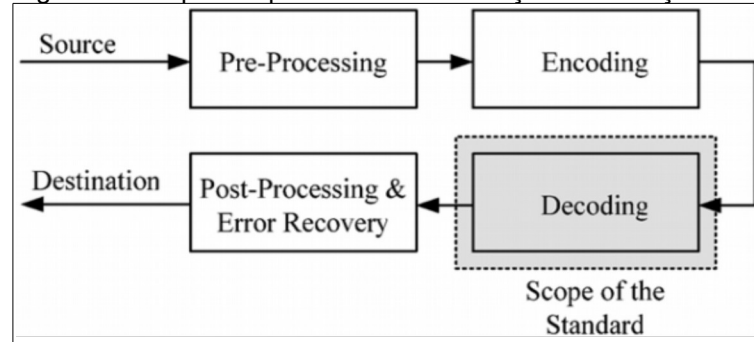


Fonte: Sullivan e Wiegand (2005, p.21).

O processo de codificação de vídeo (excluindo-se o transporte e armazenamento do sinal de vídeo) e o escopo do padrão de codificação são apresentados na figura 2. Para todos padrões de codificação de vídeo ITU-T e ISO/IEC JTC 1, apenas o decodificador central é padronizado. Em termos gerais, os padrões definem uma sintaxe específica de fluxo de dados, impondo poucas restrições sobre o valor desta sintaxe. Eles também definem um escopo específico para o processo de decodificação. A intenção é que, para cada codificador que está

de acordo com o padrão, haja uma saída similar dado certo fluxo de dados de entrada.

Figura 3 – Etapas do processo de codificação/decodificação.



Fonte: Sullivan e Wiegand (2005, p. 22).

### 3 PADRÃO H.264/AVC

A tecnologia de vídeo digital está possibilitando e, constantemente, gerando novas aplicações com um crescente aumento de requisitos em relação as características básicas de vídeo, como: resolução espacial-temporal, formato cromático e precisão de amostragem. Aplicações atuais variam desde vídeo-conferência à *mobile* TV, transmissão televisiva no modo padrão/alta-definição, aplicações de altíssima qualidade - como as de gravação de vídeo digital profissional – ou mesmo cinema digital/telas digitais amplas (MARPE; WIEGAND; SULLIVAN, 2006, tradução nossa).

Padrões de vídeo anteriores como MPEG2/H.262, H.263, e MPEG4 *Part* 2, já estão em vigor parcialmente nas áreas de aplicações citadas anteriormente. Porém, com a proliferação de vídeo digital em novos ambientes de aplicação como, *mobile* TV ou transmissão de TV em alta definição, os requisitos para uma representação eficiente aumentaram à níveis tais que tecnologias de codificação de padrões anteriores mal podem acompanhar. Além disso, mais soluções otimizadas em termos de taxa de bits vs. qualidade *end-to-end* de reprodução, estão, também, cada vez mais sendo solicitadas em áreas de aplicações tradicionais de vídeo digital (MARPE; WIEGAND; SULLIVAN, 2006, tradução nossa).

O padrão H.264/AVC, aprimorou substancialmente a robustez a erro e ampliou a flexibilidade e o escopo de aplicabilidade, comparativamente aos seus predecessores.

#### 3.1 CAMADA DE ABSTRAÇÃO DE REDE

A Camada de Abstração de Rede (NAL), estabelece o formato e informação necessárias para transporte do conteúdo do vídeo pela rede. Esta camada é projetada para permitir uma customização simples e eficiente do uso da Camada de Vídeo Codificado (VCL), para uma ampla variedade de sistemas. A completa adequação do conteúdo do vídeo de acordo com as necessidades de cada aplicação particular, está fora do escopo do padrão H.264/AVC. Porém, o desenvolvimento da NAL antecipa a variedade destas demandas (SULLIVAN; WIEGAND, 2005, tradução nossa). Alguns blocos principais da estrutura da Camada de Abstração de Rede, são: as unidades NAL, parâmetros de configuração

(*parameter sets*), e unidades de acesso (*access units*). Uma breve descrição destes conceitos é dada abaixo.

### 3.1.1 Unidades NAL

Os dados do vídeo codificado são organizados em unidades NAL, os quais são pacotes que contém um número *integer* de bytes. O primeiro byte de cada unidade NAL é um cabeçalho que contém uma indicação de qual o tipo de dado na unidade NAL, e os bytes restantes contém os dados do tipo indicado pelo cabeçalho. Existem duas classes de unidades NAL, chamadas VCL e não-VCL. As unidades NAL VCL contém dados que representam os valores das amostras das imagens do vídeo. Já as unidades NAL não-VCL, contém todo o restante da informação como, parâmetros de configuração e informação suplementar relacionadas à qualidade do vídeo.

### 3.1.2 Parâmetros de configuração

Os parâmetros de configuração, contém informações de cabeçalho importantes que podem ser aplicadas a um grande número de unidades VCL NAL. Há dois tipos de parâmetros de configuração:

- a) parâmetros de configuração sequenciais, os quais se aplicam a uma série de imagens de vídeo codificadas;
- b) parâmetros de configuração de imagens, os quais se aplicam à decodificação de uma ou mais imagens individuais.

Cada uma das principais unidades NAL VCL relacionada a uma imagem, contém um identificador que refere-se ao conteúdo do parâmetro de configuração da imagem. Cada parâmetro de configuração da imagem, por sua vez, contém um identificador que se refere ao parâmetro de configuração sequencial. Dessa forma, uma pequena quantidade de dados (o identificador) pode ser usado para estabelecer uma quantidade maior de informação (o parâmetro de configuração) sem repetir a informação dentro de cada unidade NAL VCL.

Os parâmetros de configuração sequencial e de imagem, descorrelaciona a transmissão de informações de alteração não-frequentes da transmissão de valores de amostragem das imagens do vídeo.

Parâmetros de configuração sequencial e de imagem, podem ser enviados consideravelmente antes das unidades NAL VCL ao qual eles se aplicam, assim como ser repetidos para garantir robustez contra perda de dados. Em algumas aplicações, parâmetros de configuração podem ser enviados dentro do canal que transporta as unidades NAL VCL (denominada transmissão “*in-band*”). Em outras aplicações pode ser vantajoso transmitir os parâmetros de configurações “*out-of-band*”, usando um mecanismo de transporte mais confiável.

### **3.1.3 Unidades de Acesso**

O conjunto de unidades NAL, VCL e não-VCL, que é associado com uma única imagem decodificada é chamada por unidade de acesso. A unidade de acesso contém todos os macroblocos de uma imagem, possivelmente algumas aproximações redundantes de algumas partes da imagem para fins de resiliência à erros (chamados de fatias redundantes), e outras informações suplementares associadas com a imagem.

## **3.2 CAMADA DE VÍDEO CODIFICADO**

O design da Camada de Vídeo Codificado (VCL), segue o método descrito anteriormente de codificação de vídeo híbrida, baseada em blocos (figura 1). Uma explicação mais detalhada sobre a estrutura da VCL é dada abaixo.

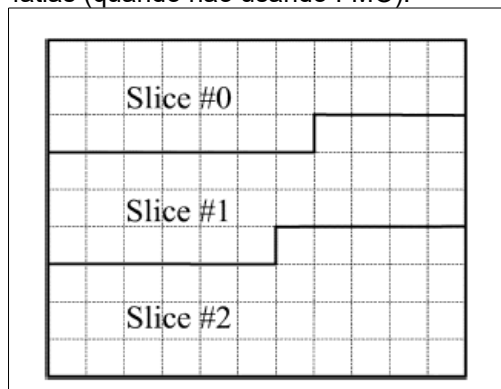
### **3.2.1 Macroblocos, Fatias, e Grupos de Fatias**

Uma sequência de vídeo codificada em H.264/AVC consiste de uma sequência de imagens codificadas. Cada imagem é particionada em macroblocos de tamanho fixo. Cada um dos macroblocos, contém uma área retangular de amostragem da imagem, de tamanho 16 pixels por 16 pixels – ou simplesmente 16x16 - para o componente de luminância e uma região de amostragem de tamanho 8x8 para cada um dos dois componentes de crominância. Macroblocos são os blocos básicos constituintes, para os quais o processo de decodificação é especificado. As amostras de luminância e crominância de um macrobloco são preditos – tanto espacialmente como temporalmente – e o resíduo da predição resultante é transmitido usando codificação por transformação. Cada componente de

cor residual é subdividida em blocos, cada bloco é transformado usando uma transformada *integer* e os coeficientes da transformada são quantizados e codificados por entropia.

Os macroblocos da imagem são organizados em fatias, as quais representam regiões de uma dada imagem que podem ser decodificadas independentemente. Cada fatia é uma sequência de macroblocos que é processada na ordem do escaneamento da rasterização, ou seja, um escaneamento do topo-esquerdo até a base-direita (embora elas não sejam necessariamente sempre consecutivas no escaneamento da rasterização, como descrito abaixo para a característica de ordenação flexível de macroblocos (FMO). Uma imagem pode conter uma ou mais fatias, como exemplificado na figura 4.

Figura 4 - Subdivisão de uma imagem em fatias (quando não usando FMO).



Fonte: Sullivan e Wiegand (2005, p.24).

Cada fatia se auto-contém, no sentido de que podem ser decodificadas sem o uso de dados de outras fatias da imagem (uma vez que todas as imagens decodificadas de referência são idênticas no codificador e decodificador para uso em MCP). No entanto, para completa decodificação, algumas informações de outras fatias podem ser necessárias para aplicar o filtro de deblocação, ao longo dos limites/arestas das fatias. Fatias podem ser usadas para:

- a) resiliência a erros, à medida que o particionamento da imagem permite um “encapsulamento” espacial, ou seja, dentro da própria imagem. Além disso, o início de cada fatia fornece um ponto de resincronização no qual o processo de decodificação pode ser reiniciado;
- b) criar pacotes que distribuam de forma otimizada os dados a serem transmitidos, respeitando o tamanho máximo de transferência da

unidade (MTU), de uma rede (por exemplo, o tamanho da MTU para *Ethernet* é de 1500 bytes);

- c) processamento paralelo, cada fatia podendo ser codificada e decodificada independentemente de outras fatias da imagem.

O aspecto da resiliência à erros das fatias pode ser aprimorada através do uso da técnica de FMO, que modifica o modo como macroblocos são associados as fatias. Usando FMO, uma imagem pode ser dividida em vários padrões de escaneamento de macroblocos. Por exemplo, fatias intercaladas é uma alocação dispersa de macroblocos, onde um ou mais grupos de fatias estão em “primeiro plano” e outro grupo de fatias “sobrando”, ou, ainda, uma tabela de mapeamento de fatias.

Uma vez que, cada fatia de uma imagem pode ser decodificada independentemente de outras, uma ordem de decodificação não é obrigatoriamente necessária. Daí surge um conceito intimamente relacionado à FMO, que diminui a robustez – que neste contexto pode ser interpretado como benéfico - e o *delay* do processo, o qual é o Ordenamento de Fatia Arbitrário (ASO). Quando o ASO está em uso, as fatias de uma imagem podem ou não estar em qualquer ordem no fluxo de dados. Quando não estão, as fatias devem estar ordenadas de modo que o primeiro macrobloco, em cada fatia subsequente, esteja na ordem da rasterização da imagem.

A perda de robustez, pode, também, ser aprimorada separando-se dados mais importantes (como tipos de macroblocos e valores de MV) de dados menos importantes (como coeficientes de valores *inter* residuais transformados), assim como, refletindo-se a importância e dependência de dados utilizando pacotes separados de unidades NAL para dados de diferentes categorias. Isso é chamado de particionamento de dados. Perca de robustez adicional pode ser alcançada enviando-se dados de codificação duplicados, de algumas ou todas as partes da imagem. Estas são chamadas de fatias redundantes.

### 3.2.2 Tipos de Fatias

Existem cinco tipos de fatias fundamentais:

- a) **fatia I**: uma fatia na qual todos os macroblocos são codificados utilizando predição *Intra*;
- b) **fatia P**: Em adição aos tipos de codificação da fatia I, macroblocos da fatia P podem também ser codificados utilizando a predição *Inter*, com no máximo um sinal MCP por bloco.
- c) **fatia B**: Em adição aos tipos de codificação disponíveis na fatia P, macroblocos de uma fatia B podem também ser codificados utilizando predição *Inter*, com dois sinais MCP por bloco, combinados usando uma média ponderada.
- d) **fatia SP**: A chamada fatia P comutável, é codificada de forma que mudanças eficientes entre diferentes transmissões de vídeo tornam-se possíveis sem o grande número de bits necessários para o caso de uma fatia I.
- e) **fatia SI**: Uma chamada fatia I comutável, que permite uma compatibilidade perfeita com uma fatia SP para acesso randômico e para fins de recuperação de erro, fazendo uso apenas da predição *Intra*.

Os três primeiros tipos de fatia listados acima são muito similares aos métodos de codificação utilizados em padrões anteriores, com exceção do uso de imagens de referência, descrito adiante.

### 3.2.3 Predição Intra de fatias

Dois tipos de codificação *Intra* de fatias são suportados: *Intra 4x4* e *Intra 16x16*. O modo *Intra 4x4* é baseado na predição de cada bloco 4x4 de luminescência separadamente, e atende as necessidades de codificação de uma imagem com quantidade significativa de detalhes. O modo *Intra 16x16*, por outro lado, faz predição e codificação residual por todo o bloco 16x16 de luminescência, sendo mais adequado para codificação de áreas de uma imagem que requerem suavidade. Adicionalmente a estes dois tipos de predição de luminescência, uma predição cromática é conduzida separadamente, sendo de tamanho de bloco fixo para os dois casos.

Em contraste com padrões anteriores de codificação de vídeo (especialmente H.263 e MPEG-4 Visual), onde predição *Intra* foi conduzida no domínio da transformação, predição *Intra* no H.264/AVC é conduzida no domínio espacial, tomando como referência amostras vizinhas de blocos previamente decodificados, que se localizam à esquerda ou acima do bloco a ser predito.

No modo *Intra* 4x4, cada bloco 4x4 de luminescência é predito a partir de amostras vizinhas espacialmente como ilustrado no lado esquerdo da figura 4. As 16 amostras do bloco 4x4, marcadas a-p, são preditas utilizando combinações de posições lineares específicas de amostras previamente decodificadas, marcadas de A-M, de blocos adjacentes. O codificador pode selecionar tanto a predição “DC” (chamada de modo 2, onde um valor médio é usado para predir o bloco inteiro) ou um de oito tipos de predição direcionais ilustrados no lado direito da figura 5. Os modos direcionais são projetados para modelar arestas de objetos de diferentes ângulos.

Figura 5 - Esquerda: predição *Intra* 4x4 é aplicada para amostras a-p utilizando amostras A-M. Direita: Oito "direções de predição" selecionáveis para modo *Intra* 4x4.



Fonte: Sullivan e Wiegand (2005, p. 25).

No modo *Intra* 16x16, todo o componente 16x16 de luminescência do macrobloco é predito de uma vez, e apenas quatro modos de predição são suportados: vertical, horizontal, DC, e plano. Os primeiros três são similares aos modos na predição *Intra* 4x4, exceto por aumentarem o número de amostras para refletir o tamanho maior do bloco. A Predição de plano usa combinações de posições lineares específicas, modelando o bloco a ser predito como um plano, através de aproximações para as arestas horizontais e verticais.

Quanto as amostras cromáticas de um macrobloco *Intra*, são preditas usando técnicas similares utilizadas para os componentes de luminosidade em macroblocos *Intra* 16x16.

### 3.2.4 Predição Inter de fatias

#### 3.2.4.1 Fatias P

Vários tipos de codificação preditivas ou de compensação de movimento são especificadas como macroblocos do tipo P. Macroblocos P podem ser particionados em regiões menores para MCP com blocos de luminosidade com tamanho de amostragem 16x16, 16x8, 8x16, e 8x8. Quando o particionamento de macrobloco 8x8 é escolhido, um elemento adicional de sintaxe é transmitido para cada partição 8x8, a qual especifica se a partição 8x8 é particionada posteriormente em regiões menores, de 8x4, 4x8, ou 4x4 para amostras de luminosidade e cromáticas correspondentes (figura 6). O sinal da predição para cada bloco de luminosidade codificado preditivamente é obtido por MC, o qual é especificado por um MV translacional e uma imagem de referência indexada. A sintaxe permite MVs indicar os limites/borda da imagem.

Figura 6 - Segmentações de macroblocos para MC. Topo: segmentação de macroblocos, Base: segmentação de partições (sub-macroblocos) 8x8.

	16x16	16x8	8x16	8x8
Macroblock (16x16) Types	0	0 1	0 1	0 1 2 3
	8x8	8x4	4x8	4x4
Sub- macroblock (8x8) Types	0	0 1	0 1	0 1 2 3

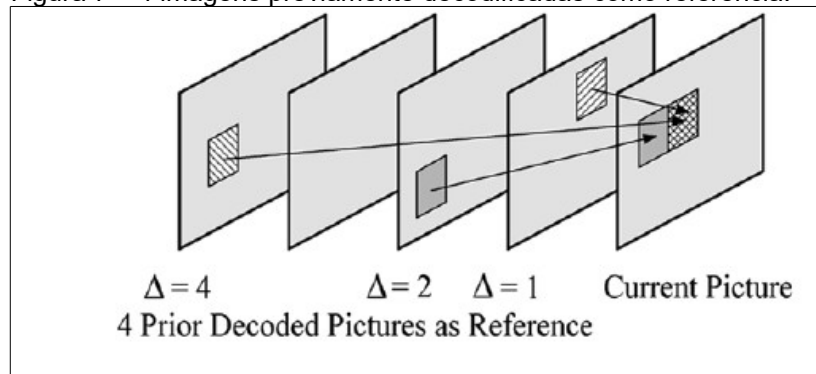
Fonte: Sullivan e Wiegand (2005, p.25).

A precisão da MC é em unidades de 1/4 da distância horizontal ou vertical entre amostras de luminescência. Se o MV aponta para uma posição *integer* de uma amostra, o sinal de predição consiste das amostras correspondentes da imagem de referência; caso contrário, a amostra correspondente é obtida usando interpolação. Os valores de predição em posições de meia-amostra são obtidos aplicando um filtro unidimensional *six-tap* de Resposta a Impulso Finito (FIR), horizontalmente e/ou verticalmente. Valores de predição em posições de 1/4 de amostra são gerados pela média de duas amostras em posições *integer* de meia-amostra.

Os valores MV são diferencialmente codificados utilizando predição mediana ou direcional de blocos vizinhos. Nenhuma predição de valor MV (ou qualquer outra forma de predição) acontece ao longo das bordas/arestas das fatias.

A sintaxe suporta MCP de multi-imagem. Ou seja, mais de uma imagem previamente decodificada pode ser usada como referência para MCP. A figura 7 ilustra o conceito. Imagens previamente decodificadas são armazenadas no Buffer de Imagens Decodificadas (DPB), encaminhadas pelo decodificador, e um índice DPB de referência é associado com cada bloco de luminosidade com movimentação compensada 16x16, 16x8, 8x16, ou 8x8. MCP para regiões menores que 8x8 usam o mesmo índice de referência para predir todos os blocos numa região 8x8.

Figura 7 – 4 imagens previamente decodificadas como referência.



Fonte: Sullivan e Wiegand (2005, p.25).

Um macrobloco P pode também ser codificado no chamado modo *P\_Skip*. Para este modo de codificação, nem uma predição quantizada de erro de sinal, nem um MC com um índice de referência são enviados. O sinal reconstruído é obtido usando apenas um sinal de predição como o aquele de um macrobloco *P\_16x16* que referencia a imagem localizada no índice 0 na lista (chamada de lista 0) de imagens no DPB. O MV usado para reconstruir o macrobloco *P\_Skip* é similar ao preditor MV para o bloco 16x16. O resultado útil deste tipo de codificação *P\_Skip* é que, amplas áreas sem mudança ou movimento constante podem ser representadas com muitos poucos bits.

#### 3.2.4.2 Predição Inter em fatias B

Em comparação a padrões de codificação anteriores, o conceito de fatias B é generalizado no H.264/AVC. Por exemplo, certas imagens podem usar outras imagens de referência contendo fatias B para MCP, dependendo do codificador indicar que a imagem B pode ser usada como referência. Portanto, a diferença substancial entre fatias B e P, é que fatias B são codificadas de uma forma na qual, alguns macroblocos ou blocos podem usar uma média ponderada de dois valores

distintos de MCP para construir o sinal de predição. Fatias B usam duas listas de imagens de referência no DPB, as quais são chamadas como primeira (lista 0) e segunda (lista 1) listas de referência, respectivamente.

Fatias B usam um particionamento de macroblocos similar à fatias P. Além de P\_16x16, P\_16x8, P\_8x16, P\_8x8, tipos de codificação *Intra*, e predição bipreditiva, outro tipo de predição chamado predição Direta (DP), é fornecido. Para cada partição 16x16, 16x8, 8x16, e 8x8, o método de predição pode ser escolhido separadamente. Uma partição 8x8 de um macrobloco B pode também ser codificada no modo Direto. Se nenhum erro de sinal de predição é transmitido para um macrobloco de modo Direto, é também chamado como modo *B Skip* e pode também ser codificado com bastante eficiência, similar ao modo *P Skip* nas fatias P. A codificação MV é similar aquela das fatias P com modificações apropriadas pelo fato de que blocos vizinhos podem ser codificados usando diferentes modos de predição.

#### 3.2.4.3 Predição ponderada em fatias P e B

Em padrões anteriores, bipredição foi tipicamente executada com uma simples média de dois sinais de predição, e a predição nos chamados tipos de macrobloco P não utilizaram ponderação. No entanto, no H.264/AVC, um codificador pode especificar certos parâmetros, escalando pesos e compensações para serem usados para cada sinal de predição nos macroblocos P e B de uma fatia. Os valores dos pesos e compensações podem ser deduzidos de relações temporais, ou, podem ser explicitamente especificados. É até mesmo permitido diferentes pesos e compensações dentro de uma mesma fatia, para execução da MCP utilizando a mesma imagem de referência.

#### 3.2.5 Transformação, Escalamento e Quantização

Similarmente aos padrões anteriores, o H.264/AVC utiliza codificação espacial por transformação do resíduo restante do processo de predição. No entanto, a transformação é aplicada à blocos de tamanho 4x4 (ao invés de blocos 8x8, tamanho utilizado por padrões anteriores). Existem diversas razões para usar transformações de tamanho menor (4x4) do que as usadas em padrões anteriores (8x8), incluindo: melhoria no decorrelacionamento espacial dado que, o resíduo de métodos cada vez mais eficazes de predição *Intra* e *Inter* tem cada vez menor

correlação espacial; uma transformada de blocos menores traz maiores benefícios visuais resultando em menos ruído ao redor de bordas/arestas; uma transformada menor requer menos computação e um comprimento de palavra menor. Ainda na codificação por transformação, é aplicada uma segunda transformada de mesmo tamanho de bloco para o modo *Intra* 16x16, visando melhorar o desempenho da compressão para regiões bastante suaves da imagem.

O parâmetro de quantização (QP) é utilizado para determinar o “grau” de quantização dos coeficientes obtidos pelo processo de transformação, podendo assumir 52 valores. O tamanho do passo de quantização é controlado logaritmicamente pelo QP ao invés de linearmente como em padrões anteriores, visando reduzir a complexidade da decodificação e aumentar o controle da taxa de bits. Cada valor do QP aumenta em 6 em relação ao anterior, duplicando o tamanho do passo de quantização. Dessa forma, cada aumento no QP aumenta o passo em aproximadamente 12% (uma mudança no tamanho do passo em ~12% significa também numa redução na taxa de bits em 12%, aproximadamente).

### 3.2.6 Codificação por entropia

No H.264/AVC, duas alternativas para codificação por entropia são suportadas. Estas são as chamadas Codificação de Comprimento Variável de Contexto-Adaptivo (CAVLC) e Codificação Binária Aritmética de Contexto-Adaptivo (CABAC). CABAC tem uma complexidade maior que CAVLC, porém, possui uma melhor eficiência de codificação.

Em ambos os modos, muitos elementos da sintaxe são codificados utilizando um único conjunto de palavra-código de comprimento infinito, chamado de código *Exp-Glomb*.

Comparado ao CAVLC, CABAC reduz a taxa de bits em 10%-15% para a mesma qualidade.

### 3.2.7 Filtro de deblocação

Uma característica incômoda da codificação baseada em blocos é a produção de um efeito observável na figura 8, especialmente em taxa de bits baixas. Bordas/arestas dos blocos são preditas por MCP com uma precisão inferior das amostras internas do bloco, e a transformação de blocos também poduz

descontinuidades nas bordas de um bloco. O efeito bloco é geralmente tido como um dos efeitos mais visíveis nos métodos de compressão atuais. Por essa razão, H.264/AVC define uma etapa de filtro de deblocação adaptivo.

O filtro reduz o efeito bloco enquanto mantém as bordas/arestas verdadeiras da cena. Conseqüentemente, a qualidade subjetiva é significativamente melhorada. O filtro geralmente reduz a taxa de bits entre 5%-10% para um determinado nível de qualidade em comparação a um vídeo sem filtragem.

Figura 8 – Desempenho do filtro de deblocação para imagens altamente comprimidas. Esq.: Sem filtragem; Dir.: Com filtragem.



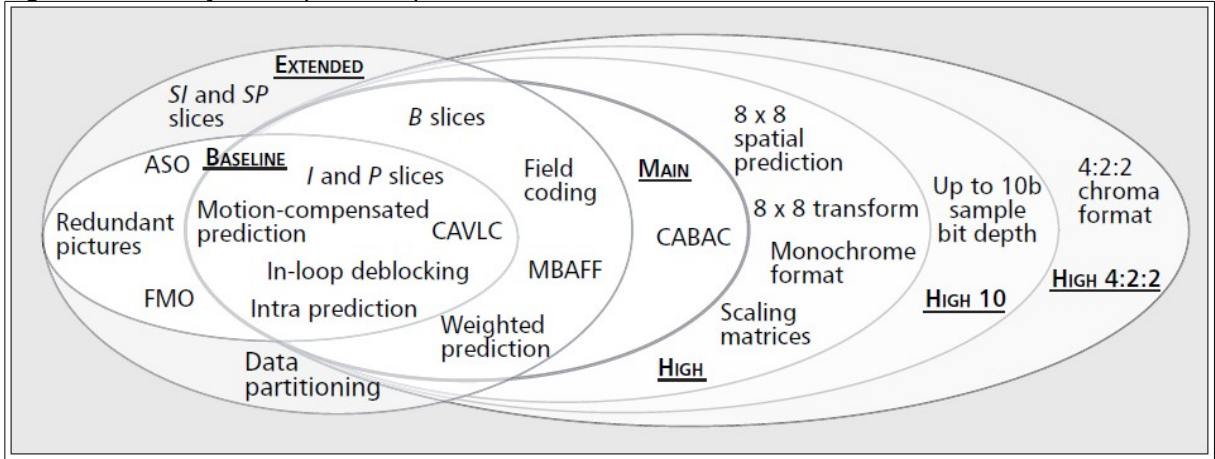
Fonte: Sullivan e Wiegand (2005, p.28).

### 3.2.8 Perfis e níveis

Perfis e níveis especificam aspectos de conformidade que garantem a interoperabilidade entre implementações do codificador e decodificador dentro de aplicações do padrão e entre várias aplicações que possuem requisitos funcionais semelhantes. Um perfil defini um conjunto de funcionalidades da sintaxe para uso na geração de um fluxo de dados adequado, de forma que um nível impõe certas restrições como, taxa de bits e tamanho da imagem máximos. Todos os decodificadores em conformidade com um perfil específico e um nível, devem suportar todas as funcionalidades incluídas naquele perfil, como especificado pelo nível. Codificadores não são obrigados a utilizarem efetivamente nenhum conjunto de funcionalidades oferecidos por um perfil e nível específico, porém, não devem violar o conjunto de funcionalidades da sintaxe nem as restrições impostas. Desta forma, é implícito que a conformidade com qualquer perfil e nível, embora garanta interoperabilidade entre decodificadores, não oferece nenhuma garantia de qualidade na reprodução *end-to-end*(MARPE; WIEGAND, 2006, tradução nossa).

A figura 9 ilustra os atuais seis perfis e suas respectivas funcionalidades, como discorrido abaixo.

Figura 9 – Ilustração dos perfis do padrão H.264/AVC.



Fonte: Marpe e Wiegand (2006, p.140).

A figura 10 apresenta os níveis do padrão H.264/AVC.

Figura 10 – Níveis do padrão H.264/AVC.

Level number	Max macroblock processing rate MaxMBPS (MB/s)	Max frame size MaxFS (MBs)	Max decoded picture buffer size MaxDpbMbs (MBs)	Max video bit rate MaxBR (1000 bits/s, 1200 bits/s, cpbBrVclFactor bits/s, or cpbBrNalFactor bits/s)	Max CPB size MaxCPB (1000 bits, 1200 bits, cpbBrVclFactor bits, or cpbBrNalFactor bits)	Vertical MV component range MaxVmvR (luma frame samples)	Min compression ratio MinCR	Max number of motion vectors per two consecutive MBs MaxMvsPer2Mb
1	1 485	99	396	64	175	[-64,+63.75]	2	-
1b	1 485	99	396	128	350	[-64,+63.75]	2	-
1.1	3 000	396	900	192	500	[-128,+127.75]	2	-
1.2	6 000	396	2 376	384	1 000	[-128,+127.75]	2	-
1.3	11 880	396	2 376	768	2 000	[-128,+127.75]	2	-
2	11 880	396	2 376	2 000	2 000	[-128,+127.75]	2	-
2.1	19 800	792	4 752	4 000	4 000	[-256,+255.75]	2	-
2.2	20 250	1 620	8 100	4 000	4 000	[-256,+255.75]	2	-
3	40 500	1 620	8 100	10 000	10 000	[-256,+255.75]	2	32
3.1	108 000	3 600	18 000	14 000	14 000	[-512,+511.75]	4	16
3.2	216 000	5 120	20 480	20 000	20 000	[-512,+511.75]	4	16
4	245 760	8 192	32 768	20 000	25 000	[-512,+511.75]	4	16
4.1	245 760	8 192	32 768	50 000	62 500	[-512,+511.75]	2	16
4.2	522 240	8 704	34 816	50 000	62 500	[-512,+511.75]	2	16
5	589 824	22 080	110 400	135 000	135 000	[-512,+511.75]	2	16
5.1	983 040	36 864	184 320	240 000	240 000	[-512,+511.75]	2	16
5.2	2 073 600	36 864	184 320	240 000	240 000	[-512,+511.75]	2	16

Fonte: Recommendation ITU-T H.264 (2014, p. 291).

### 3.2.8.1 Perfis: Baseline, Main e Extended (Versão 1)

Na primeira versão do padrão H.264/AVC, três perfis foram definidos: o perfil *Baseline*, *Extended*, e *Main*. O perfil *Baseline* suporta todas as funcionalidades especificadas pelo padrão, exceto pelos três sets seguintes:

- set 1: fatias; codificação de campo da imagem e MBAFF (para mais informações, vide Sullivan e Wiegand, 2005); predição ponderada;
- set 2: CABAC;
- set 3: fatias SP e SI; particionamento de dados de fatias.

O primeiro e segundo destes três sets é suportado pelo perfil *Main* (MP) adicionalmente as funcionalidades do perfil *Baseline*, exceto pela funcionalidade FMO e algumas outras relacionadas à resiliência a erros. O perfil *Extended* suporta

todas as funcionalidades do perfil *Baseline*, além do primeiro e terceiro sets de funcionalidades, exceto CABAC (MARPE; WIEGAND, 2006, tradução nossa).

Em termos gerais, o perfil *Baseline* foi projetado para aplicações em que um mínimo de complexidade computacional e um máximo de robustez ao erro é necessário, visto que o perfil *Main* foi projetado para aplicações que necessitam de uma eficiência de codificação máxima, com um pouco menos de ênfase em robustez ao erro. O perfil *Extended* foi projetado para ser uma junção entre os recursos dos perfis *Baseline* e *Main* com foco adicional em necessidades específicas de aplicações de *streaming* de vídeo, e futura mais robustez contra erros e perda de pacotes (MARPE; WIEGAND, 2006, tradução nossa).

### 3.2.8.2 Perfis *High* da emenda FRExt (Versão 3)

Como ilustrado na figura 8, a emenda FRExt, do inglês *Fidelity Range Extensions*, do atual padrão, especifica, adicionalmente aos três perfis da versão 1, um grupo de três perfis construídos como um conjunto de recursos sobre o perfil *Main*. O perfil *High* (HP), em sua intersecção com o perfil anteriormente mencionado, contém as ferramentas FRExt mais relevantes para futura melhoria da eficiência de codificação. Relativamente ao perfil *Main*, essas ferramentas implicam apenas numa moderada (se alguma) melhoria em complexidade em termos de implementação e custo computacional (no que se refere ao decodificador) (até 4:2:2) (MARPE; WIEGAND, 2006, tradução nossa).

Portanto, o perfil *High*, com sua restrição de vídeos 8 bits no formato cromático 4:2:0, ultrapassou o perfil *Main* na perspectiva das aplicações comerciais SD e HD. Dois outros perfis, chamados perfis *High 10* e *High 4:2:2*, expandem ainda mais a capacidade do padrão de incluir aplicações que requerem alta precisão de amostragem (até 10 bits/amostra) e formatos cromáticos superiores (até 4:2:2) (MARPE; WIEGAND, 2006, tradução nossa).

### 3.3 ESCALABILIDADE

Esta seção desenvolve uma explanação de forma genérica sobre o conceito de escalabilidade – subitem 3.3.1 - e sobre a própria extensão de Codificação Escalável de Vídeo (SVC), uma *Fidelity Range Extension* (FRExt) do padrão H.264/AVC, que visa trazer novas funcionalidades ao padrão “base”<sup>2</sup>.

#### 3.3.1 Vídeos escaláveis

Basicamente, uma sequência/fluxo binária de um vídeo é chamado escalável quando partes desta sequência podem ser removidos de uma forma que a sequência resultante forma uma outra sequência de bits válida para um determinado decodificador. A subsequência, por sua vez, representa a fonte do conteúdo com uma qualidade de reconstrução que é menor que a qualidade daquela sequência completa de bits, considerada de alta qualidade, no entanto, levando em conta a quantidade de dados residual. Sequências de bits que não fornecem esta propriedade são chamadas como sequências de bits de única camada.

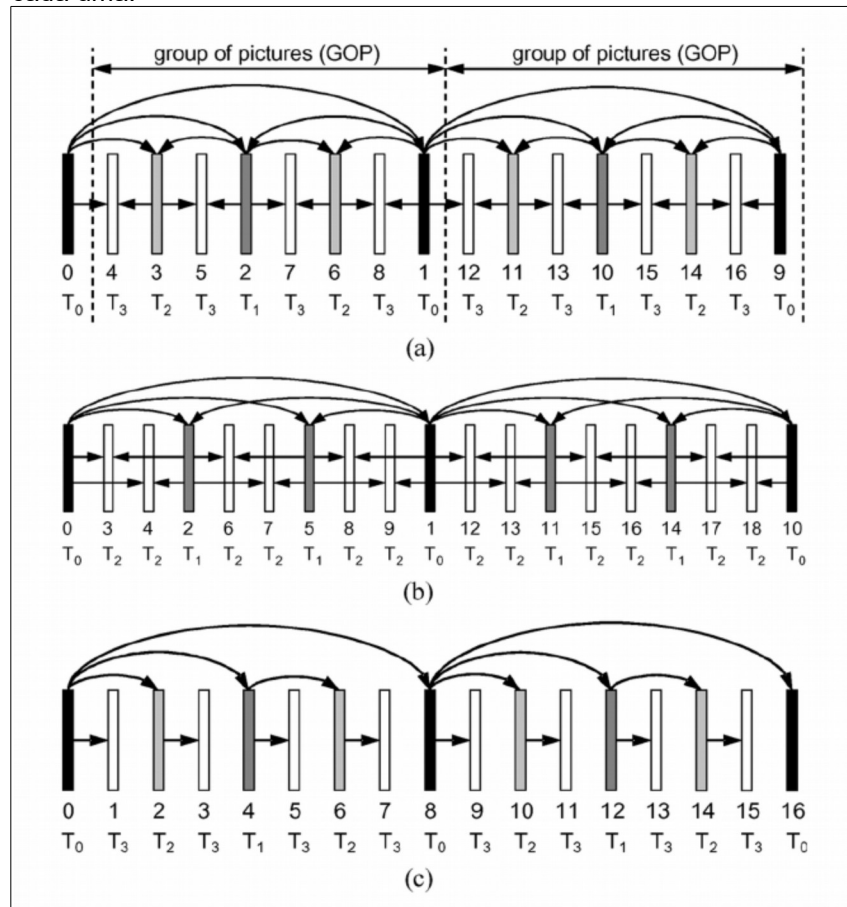
Os modos escaláveis mais comuns, são: temporal, espacial, e escalabilidade qualitativa. A Escalabilidade espacial e temporal descrevem casos nos quais subconjuntos da sequência de bits representam o conteúdo original com um tamanho de imagem reduzido (resolução espacial) ou taxa de quadros (resolução temporal), respectivamente. Já para a escalabilidade qualitativa, o subconjunto fornece a mesma resolução espacial-temporal como da sequência completa de bits, mas com uma fidelidade menor – onde fidelidade é, normalmente, informalmente associada à taxa de sinal-ruído (SNR). Escalabilidade qualitativa é, também, comumente chamada de fidelidade ou escalabilidade SNR.

Os diferentes tipos de escalabilidade podem ser combinados, de forma que numa única sequência de bits diferentes representações espaciais-temporais e taxa de bits são suportadas. A seguir é ilustrada a estrutura da escalabilidade temporal extensão SVC, figuras 11.

---

<sup>2</sup> Ao longo de toda esta subseção foi utilizada a referência Schwarz, Marpe e Wiegand (2007). Para mais informações a respeito da extensão SVC, vide o artigo citado.

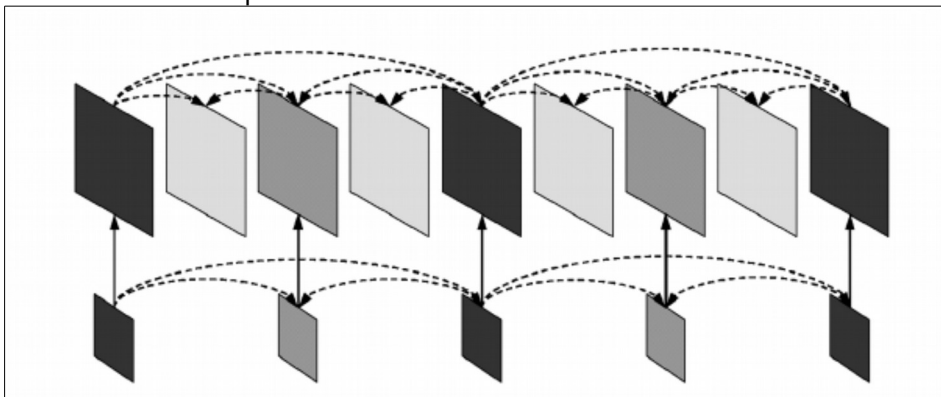
Figura 11 – Diferentes tipos de estrutura de predição hierárquica utilizada na escalabilidade temporal. Os números abaixo de cada imagem especificam a ordem de codificação e, os símbolos  $T(k)$  especificam as camadas temporais com  $k$  como identificadores de cada uma.



Fonte: Schwarz, Marpe e Wiegand (2007, p.1107).

A seguir é ilustrada a estrutura da escalabilidade espacial da extensão SVC, figura 12.

Figura 12 – Estrutura multicamada com camada intercalada adicional utilizada na escalabilidade espacial.



Fonte: Schwarz, Marpe e Wiegand (2007, p.1109).

### 3.3.2 A extensão SVC

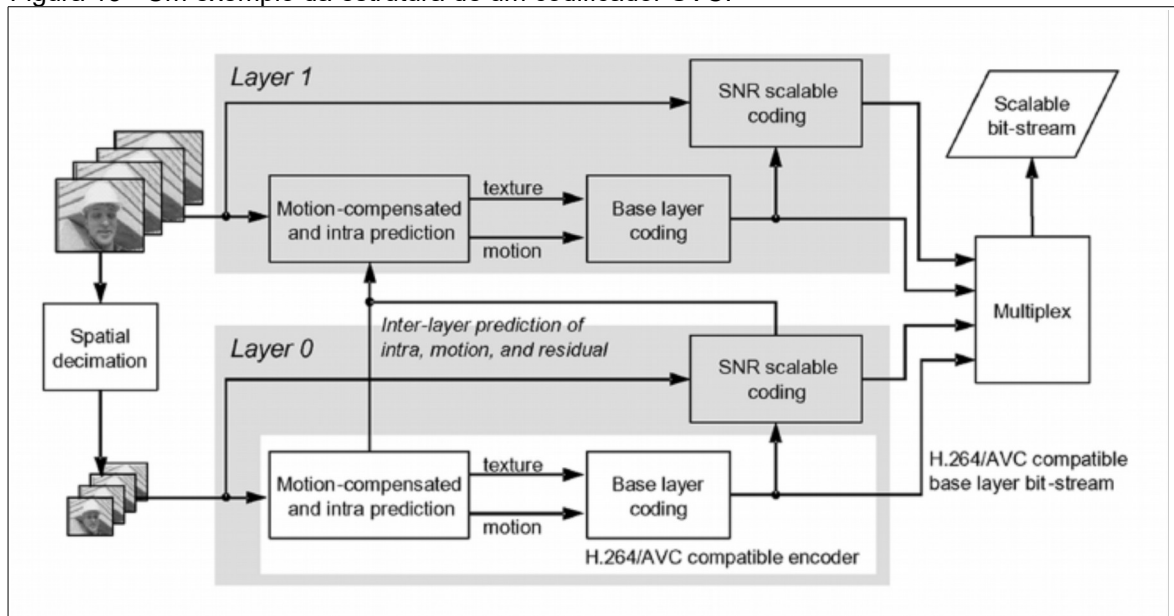
Consideremos, por exemplo, um cenário em que um serviço de transmissão de um vídeo com uma clientela heterogênea, onde múltiplas sequências de bits do mesmo conteúdo de origem, diferindo-se em tamanho da imagem codificada, taxa de quadros, e taxa de bits, devem ser transmitidos simultaneamente. Com a aplicação de um recurso escalável bem configurado, o conteúdo original tem de ser codificado apenas uma vez – para a resolução máxima e taxa de bits requisitada, resultando numa sequência de bits nas quais representações com resoluções e/ou qualidades menores, podem ser obtidas descartando dados selecionados.

Outro benefício do uso da extensão SVC é que, uma sequência de bits escalável geralmente contém partes com diferente importância em termos da qualidade do vídeo decodificado. A perda de importantes unidades de transmissão, devido à congestão, pode ser evitada e a robustez a erro total do serviço de transmissão de vídeo pode ser substancialmente melhorada.

A extensão SVC é de grande utilidade para aplicações de fiscalização, as quais exigem serem visualizadas em múltiplos aparelhos, desde monitores de alta definição até *smartphones* ou PDAs, além de serem arquivadas e armazenadas. Sequências de bits de alta resolução/alta qualidade geralmente podem ser deletadas após certa data de expiração, de forma que apenas cópias do vídeo de baixa qualidade são mantidas para arquivamento de longo prazo.

Apesar da extensão SVC oferecer uma variedade de funcionalidades de grande utilidade, os perfis escaláveis nos padrões existentes foram raramente utilizados no passado, principalmente pelo fato de a escalabilidade espacial e qualitativa serem um recurso custoso em termos de complexidade para o decodificador e da diminuição significativa na eficiência de codificação. Em contraste, escalabilidade temporal é frequentemente suportada, por exemplo, em aplicações baseadas no padrão H.264/AVC, principalmente por que este veio com um aumento de eficiência de codificação considerável. A seguir, a ilustração da estrutura de um codificador SVC na figura 13.

Figura 13 - Um exemplo da estrutura de um codificador SVC.



Fonte: Schwarz, Marpe e Wiegand (2007, p.1116).

## 4 PADRÃO H.265 HEVC

Como têm sido o caso para todos os padrões anteriores da ITU-T e ISO/IEC, apenas a estrutura do fluxo de dados e a sintaxe são padronizados, assim como, as restrições para o fluxo de dados e o seu mapeamento para geração de imagens decodificadas. Esta limitação de escopo do padrão, permite liberdade máxima para a otimização de implementações, de acordo com a especificidade de cada aplicação (balanceamento de compressão, custo de implementação, tempo de desenvolvimento, e outras variáveis). No entanto, o padrão por si só não provê nenhuma garantia de qualidade de reprodução *end-to-end* (SULLIVAN et al, 2012, tradução nossa).

O padrão H.265/HEVC foi publicado em 2013 e basicamente mantém a estrutura desenvolvida no padrão anterior H.264/AVC, trazendo novos recursos que aprimoram o processo de codificação/decodificação. Ao longo deste capítulo será explanado em termos gerais as mudanças estruturais e novas abordagens advindas<sup>3</sup>.

### 4.1 DESIGN E RECURSOS

As seguintes subseções descrevem brevemente os elementos principais do padrão, assim como seu funcionamento. Observa-se que foi usado como base a estrutura do padrão anterior H.264/AVC. Sendo assim, certos conceitos permanecem, já descritos no capítulo anterior.

Note-se que foi mantida a terminologia em inglês na maior parte do texto, tendo em vista evitar possíveis confusões e da praticidade descritiva.

#### 4.1.1 Camada de Codificação de Vídeo (VCL)

Assim como no padrão H.264/AVC, o padrão H.265/HEVC possui duas camadas, uma responsável pela formatação e transmissão do fluxo de bits interno do codificador e outra responsável unicamente pela codificação, a qual denomina-se Camada de Codificação de Vídeo (VCL).

---

<sup>3</sup> A referência Sullivan et al (2012) foi extensivamente utilizada como base deste capítulo, sendo referenciado no decorrer do texto apenas referências adicionais. Para um aprofundamento maior, favor consultar a bibliografia utilizada neste trabalho, que inclui a norma do padrão H.265/HEVC.

Para o padrão H.265/HEVC, houve um aumento estrutural considerável que trouxe a implementação de novas estruturas de dados e amostragem. Nos subitens subsequentes serão descritas as mudanças implementadas durante o processo de codificação.

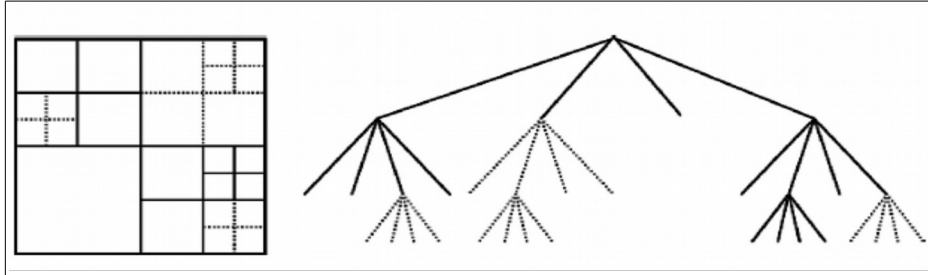
#### 4.1.1.1 Coding Tree Units, Coding Tree Blocks, Coding Units, e Coding Blocks

O núcleo da camada de codificação, em padrões passados, era o macrobloco – sendo este constituído por um bloco 16x16 de amostras de luminescência, e, dois blocos correspondentes 8x8 de amostras de crominância, para uma amostragem de cores 4:2:0 – , ao passo que, a estrutura análoga no padrão H.265/HEVC é a *Coding Tree Unit* (CTU), a qual possui uma estrutura de árvore quaternária podendo ser desmembrada em unidades menores – discutidas a seguir -, e que possuem um tamanho definido pelo codificador, podendo assumir tamanhos maiores que um macrobloco.

A CTU consiste dos *Coding Tree Blocks* (CTBs) de luminescência e cromáticos correspondentes, além dos elementos de sintaxe. O tamanho  $L \times L$  de um CTB de luminescência pode assumir valores  $L = 16, 32$  ou  $64$  amostras, com um aumento da eficácia de compressão à medida que o tamanho aumenta. A raiz da árvore quaternária está diretamente relacionada com a CTU.

Os CTBs, por sua vez, também podem ser particionados em blocos menores, seguindo a mesma estrutura de árvore quaternária. Tais blocos menores são chamados de *Coding Blocks* (CBs). A sintaxe de árvore quaternária da CTU especifica o tamanho e posição dos *Coding Blocks* (CBs) de luminescência e de crominância, ilustrado na figura 14. A divisão de um CTU em CBs de luminescência e crominância, com a sintaxe associada, formam uma *Coding Unit* (CU). Um CTB pode conter apenas uma CU, ou, pode ser dividido para formar múltiplas CUs. A cada CU, associa-se mais uma segmentação, desta vez em *Prediction Units* (PUs) – as quais irão se encarregar do processo de predição *Intra* ou *Inter* - e em *Transform Units* (TUs) – as quais são ramificadas em *Transform Blocks* (TBs), que são blocos residuais do processo de predição transformados por uma transformada -, explanadas nas próximas subseções.

Figura 14 – Subdivisão de um CTB em CBs (e *Transform Blocks* - TBs). Linhas contínuas indicam as arestas dos CBs; linhas pontilhadas indicam arestas de TBs (Lado esq. Particionamento do CTB; Lado dir. *quadtree* correspondente).



Fonte: Sullivan et al (2012, p. 1656).

#### 4.1.1.2 Prediction Units e Prediction Blocks

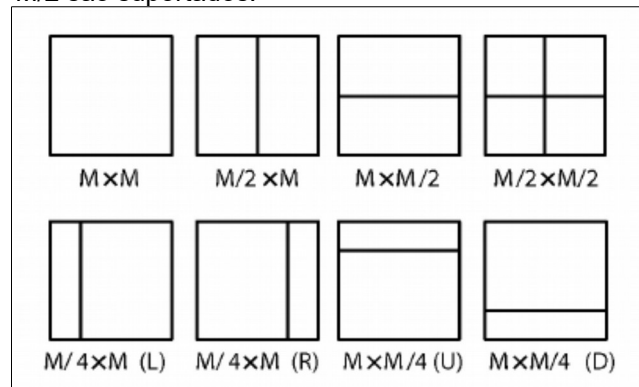
Os *Prediction Blocks* (PBs), juntos com a sintaxe de predição, formam a *Prediction Unit* (PU).

A decisão de codificar a área de uma imagem ou não, utilizando predição *Inter* ou *Intra*, é feita a nível da CU. Sendo assim, o particionamento da PU tem sua raiz no nível da CU. Dependendo da decisão de qual tipo de predição a ser empregado, os CBs de luminescência e crominância podem ser posteriormente divididos em, e preditos, a partir de PBs, de luminescência e de crominância. O presente padrão suporta tamanhos variáveis de PBs, desde amostras 4x4 até 64x64.

Quando o modo de predição é definido como *Intra*, o tamanho do PB, o qual é o tamanho do bloco no qual o modo predição *Intra* é estabelecido, é igual ao tamanho do CB para todos os blocos, exceto o menor CB permitido no fluxo de dados.

Quando o modo de predição é definido como *Inter*, é especificado se os CBs luminosos e cromáticos são divididos em um, dois ou quatro PBs (figura 15).

Figura 15 - Modos para divisão de um CB em PBs, sujeita a certas limitações de tamanho. Para CBs preditos por predição *Intra*, apenas  $M \times M$  e  $M/2 \times M/2$  são suportados.



Fonte: Sullivan et al (2012, p.1655).

#### 4.1.1.3 Transform Units e Transform Blocks

Uma *Transform Unit* (TU) é uma ramificação partindo de um CB que, segue a estrutura de árvore quaternária da CTU. Seu nó correspondente está a nível da CU, sendo particionado em *Transform Blocks* (TBs).

Portanto, para a codificação residual, um CB pode ser transformado por inteiro, ou, recursivamente particionado em *Transform Blocks* (TBs). O particionamento é representado por uma árvore quaternária residual (veja a figura 14). Os blocos correspondentes aos nós-folha, resultantes da *quadtree* residual (TBs), são processados por codificação por transformação.

Funções *integer* semelhantes aquela da Transformada Discreta do Cosseno (DCT) são definidas para os TBs de tamanho  $4 \times 4$ ,  $8 \times 8$ ,  $16 \times 16$ , e  $32 \times 32$ . Para a transformação  $4 \times 4$  de resíduos da predição de luminescência, uma transformada *integer* derivada da Transformada Discreta do Seno (DST) é fornecida alternativamente.

#### 4.1.1.4 Representação de Vetores de Movimento

A predição Avançada de Vetor de Movimento (AMVP), é empregada, baseando-se em dados de PBs adjacentes e a imagem de referência. Um modo de mesclagem para codificação de MVs também pode ser utilizado, permitindo a herança de MVs de PBs, temporalmente ou espacialmente, vizinhos. Além disso,

assim como no padrão H.264/AVC, a inferência de movimento *skipped* ou *direct* (modos *Skip* e *Direct* de predição) também é disponibilizada.

#### 4.1.1.5 Compensação de Movimento

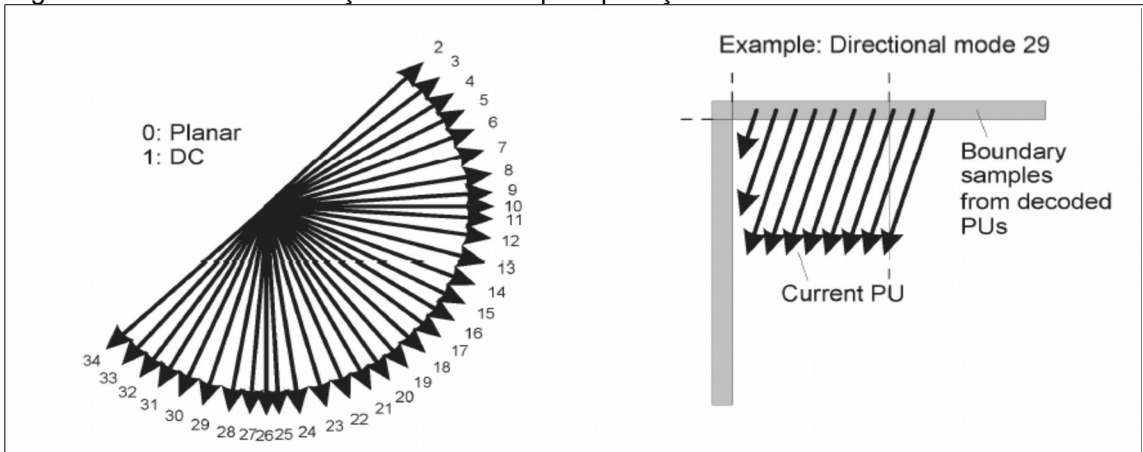
Precisão de um quarto de amostra é utilizada para os MVs, e filtros *7-tap* e *8-tap* são utilizados para interpolação de posições de amostras fracionárias (comparado ao filtro *6-tap* de posições de meia-amostra seguidos por interpolação linear para posições de um quarto de amostra no H.264/AVC). Similarmente ao H.264/AVC, múltiplas imagens de referência são utilizadas. Para cada PB, um ou dois vetores de movimento podem ser transmitidos, resultando numa codificação unipreditiva ou bipreditiva, respectivamente. Assim como no H.264/AVC, escalamento e operação de balanceamento podem ser aplicados aos sinais de predição, métodos conhecidos como predição ponderada.

#### 4.1.1.6 Predição *Intra*

A predição *Intra* opera de acordo com o tamanho do TB e arestas de blocos previamente decodificados adjacentes de TBs espacialmente vizinhos que são usados para formar o sinal de predição. A predição direcional com 33 orientações direcionais diferentes é definida para tamanhos de TB desde 4x4 até 32x32. As possíveis direções de predição são ilustradas na figura 16. Alternativamente, predição planar e predição DC podem ser empregadas.

Cada CB pode ser codificado por um ou mais tipos de codificação, dependendo do tipo da fatia. Similarmente ao padrão H.264/AVC, a codificação preditiva *Intra* é suportada para todos os tipos de fatias. O presente padrão suporta vários métodos de codificação preditiva *Intra*, chamados de Intra-Angular, Intra-Planar e Intra-DC.

Figura 16 – Modos e orientações direcionais para predição *Intra*.



Fonte: Sullivan et al (2012, p.1657).

#### 4.1.1.7 Controle de quantização

Assim como no H.264/AVC, a reconstrução uniforme por quantização (URQ) é utilizada no H.265/HEVC, com matrizes escaláveis de quantização suportadas para os vários tamanhos de blocos transformados.

#### 4.1.1.8 Codificação por entropia

A codificação CABAC é utilizada para a entropia, semelhantemente no padrão H.264/AVC, porém foi submetida a diversas melhorias no que se refere à velocidade de transmissão (especialmente para arquiteturas de processamento paralelo), seu desempenho de compressão, e para redução dos requisitos de memória.

#### 4.1.1.9 Filtro de deblocação *in-loop* (aninhado em laço)

Um filtro de deblocação semelhante ao utilizado no H.264/AVC é executado dentro do laço da predição *Inter*. No entanto, o design é simplificado em relação a sua tomada de decisão e processos de filtragem, além de possuir maior compatibilidade com processamento paralelo.

#### 4.1.1.10 Perfis e níveis

Foram projetados apenas três perfis para diferentes requisitos de aplicações, chamados de perfis *Main*, *Main 10*, e *Main Still Picture*. Diminuir o número de perfis fornece uma interoperabilidade máxima entre dispositivos, e é reafirmada pelo fato de que serviços tradicionais distintos, como *broadcast*, *mobile*, *streaming*, estão convergindo para o ponto onde a maioria dos dispositivos devem se tornar capazes de suportar todos eles. Os três perfis consistem de ferramentas de codificação e sintaxe de alto nível, enquanto impõem as seguintes restrições:

- a) apenas sub-amostragem cromática 4:2:0 é suportada;
- b) quando um codificador codifica uma imagem utilizando *tiles* múltiplos, este não pode utilizar, também, processamento de frente-de-onda paralelo, e cada *tile* deve ter ao menos 256 amostras de luminância de comprimento e 64 amostras de luminância de altura;
- c) nos perfis *Main* e *Main Still Picture*, apenas uma precisão de 8 bits por amostra é suportada, enquanto que o perfil *Main 10* suporta até 10 bits por amostra;
- d) no perfil *Main Still Picture*, o fluxo de dados completo deve conter apenas uma imagem codificada (e portanto, predição *Inter* não é suportada). Atualmente, a definição de 13 níveis, como apresentados na figura 17, variando desde níveis que suportam apenas imagens relativamente pequenas em tamanho, como tamanho de imagem de luminância de 176x144 pixels até imagens de tamanho igual a 7680x4320 (geralmente chamadas de 8k x 4k).

Figura 17 – Níveis do padrão H.265/HEVC

Level	Max luma sample rate MaxLumaSr (samples/sec)	Max bit rate MaxBR (BrVcFactor or BrNcFactor bits/s)		Min compression ratio MinCrBase	
		Main tier	High tier	Main tier	High tier
1	552 960	128	-	2	2
2	3 686 400	1 500	-	2	2
2.1	7 372 800	3 000	-	2	2
3	16 588 800	6 000	-	2	2
3.1	33 177 600	10 000	-	2	2
4	66 846 720	12 000	30 000	4	4
4.1	133 693 440	20 000	50 000	4	4
5	267 386 880	25 000	100 000	6	4
5.1	534 773 760	40 000	160 000	8	4
5.2	1 069 547 520	60 000	240 000	8	4
6	1 069 547 520	60 000	240 000	8	4
6.1	2 139 095 040	120 000	480 000	8	4
6.2	4 278 190 080	240 000	800 000	6	4

Fonte: Recommendation ITU-T H.265 (2014, p. 235)

#### 4.1.2 Camada de Abstração de Rede (NAL)

Diversos aspectos novos de design no presente padrão aumentam a flexibilidade para operação sobre uma diversidade de aplicações e ambientes de rede, além de aprimorarem a robustez em relação à perda de dados. No entanto, a sintaxe de alto nível utilizada no padrão H.264/AVC foi, de modo geral, mantida, incluindo os seguintes recursos.

##### 4.1.2.1 Estrutura dos parâmetros de configuração

Parâmetros de configuração contém informação que pode ser compartilhada para a decodificação de diversas regiões do vídeo decodificado. A estrutura do parâmetro de configuração fornece um mecanismo robusto para transmitir dados que são essenciais ao processo de decodificação.

#### 4.1.2.2 Estrutura da sintaxe da unidade NAL

Cada estrutura de sintaxe é colocada em um pacote lógico de dados chamado de unidade NAL. Utilizando o conteúdo de 2 bytes do cabeçalho de uma unidade NAL, é possível de imediato identificar o conteúdo dos dados “carregados”.

#### 4.1.2.3 Fatias

Uma fatia é uma estrutura de dado que pode ser decodificado independentemente de outras fatias da mesma imagem, em termos de codificação por entropia, predição de sinal, e reconstrução de sinal residual. Uma fatia pode ser tanto uma imagem inteira ou uma região de uma imagem. Um dos principais propósitos das fatias é a resincronização durante a perda de dados. No caso de transmissão por pacotes, o número máximo de bits carregados dentro de uma fatia é, geralmente, restrito. O número de CTUs numa fatia é frequentemente variado para minimizar sobrecarga de pacotes, mantendo-se o tamanho de cada pacote dentro do limite estabelecido (figura 18 (a)).

#### 4.1.2.4 SEI e VUI

A sintaxe inclui suporte para vários tipos de metadados conhecidos como Informação Suplementar de Aperfeiçoamento (SEI), e Informação de Usabilidade do Vídeo (VUI). Estes dados fornecem informação sobre sincronização de imagens de vídeos, interpretação apropriada do espaço de cores utilizado no sinal do vídeo, informações para exibição, entre outras.

### **4.1.3 Elementos de arquitetura de processamento paralelo**

Quatro novos recursos são introduzidos no presente padrão para aumentar a capacidade de processamento paralelo, ou modificar a estruturação dos dados das fatias para fins de empacotamento. Cada um deles podem ter benefícios para aplicações específicas, dependendo do implementador do codificador/decodificador determinar, se e como, tirar proveito destes recursos.

#### 4.1.3.1 Tiles

A opção de particionar a imagem em regiões retangulares chamadas de *tiles* foi criada. O principal propósito dos *tiles* está mais relacionado a aumentar a capacidade de processamento paralelo do que fornecer resiliência à erro. *Tiles* são regiões independentemente decodificadas de uma imagem, que possuem algumas informações de cabeçalho compartilhadas. *Tiles* podem ser adicionalmente utilizados para acesso espacial randômico à regiões de uma imagem de um vídeo.

A configuração padrão de *tiles* de uma imagem, consiste em particionar a imagem em regiões retangulares com números aproximadamente iguais de CTUs em cada *tile* (figura 18 (b)). *Tiles* fornecem paralelismo num nível de mais baixa granularidade (imagem/subimagem) e, para o seu uso, não se faz necessária qualquer sincronização sofisticada de *threads*.

#### 4.1.3.2 Processamento de frente-de-onda

Quando o processamento paralelo de frente-de-onda (WPP), é ativado, uma fatia é dividida em linhas de CTUs. A primeira linha é processada de forma padrão. A segunda linha pode começar a ser processada apenas depois de dois CTUs da primeira terem sido processados. A mesma lógica se aplica à terceira e demais linhas (figura 18 (c)).

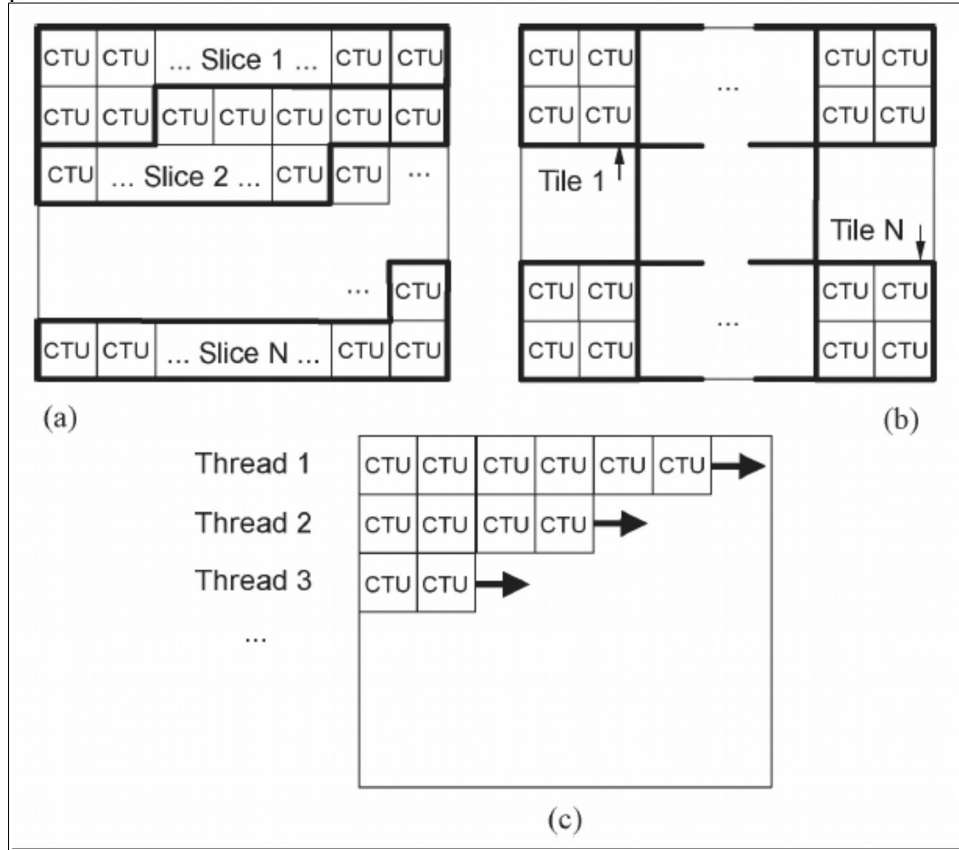
O WPP fornece uma forma de processamento paralelo num nível alto de granularidade. Ele normalmente oferece melhor desempenho de compressão que *tiles* (e evita alguns incômodos visuais induzidos pelos mesmos, distúrbios na imagem chamados em inglês de *visual artifacts*).

#### 4.1.3.3 Segmentos de fatia dependente

Segmentos de fatia dependente permitem que, dados associados com um ponto de entrada de frente-de-onda específica ou *tile*, ser transmitido numa unidade NAL separada. Dessa forma, torna-se estes dados disponíveis mais rapidamente para um sistema de empacotamento fragmentado, resultando numa latência menor do que a codificação completa em uma única fatia. Um segmento de fatia dependente para um ponto de entrada de frente-de-onda pode apenas ser decodificado após, pelo menos, parte do processo de decodificação de outro segmento de fatia ter sido executado. Segmentos de fatia dependente são

principalmente úteis em codificação de baixa latência, onde outras ferramentas paralelas podem penalizar o desempenho da compressão.

Figura 18 – Subdivisão de uma imagem em (a) fatias, (b) *tiles*, (c) Ilustração do processamento de frente de onda.



Fonte: Sullivan et al (2012, p.1656).

## 5 TRABALHOS CORRELATOS

Neste capítulo foram listados trabalhos anteriormente conduzidos em codificação de vídeo, os quais tratam de análises e estudos de caso sobre os padrões H.265/HEVC e/ou H.264/AVC. Desta forma, procura-se demonstrar o material já existente no meio científico relacionado a presente pesquisa, o qual foi útil para o embasamento e desenvolvimento desta pesquisa.

### 5.1 ANÁLISE COMPARATIVA ENTRE OS CODIFICADORES DE VÍDEO H.264/AVC E H.265/HEVC

Em sua monografia para conclusão de curso, a autora Silva (2014), do curso de Engenharia de Telecomunicações da Escola Politécnica da Universidade de Pernambuco, procura fazer uma análise comparativa entre o padrão H.264/AVC e o novo padrão H.265/HEVC. Tem-se como intenção avaliar o desempenho de ambos os codificadores, procurando comprovar a superioridade do padrão H.265/HEVC.

Primeiramente faz-se uma explanação sobre os conceitos básicos de vídeo digital e o histórico de cada codificação de vídeo ao longo do tempo. Posteriormente, é apresentada as tecnologias utilizadas no padrão H.264.AVC, seguido pelo seu sucessor H.265/HEVC.

Discute-se, também, as métricas de avaliação de codificações de vídeo, abrangendo métodos utilizados para avaliação subjetiva da qualidade dos vídeos codificados.

A monografia parte da premissa, assim como o presente trabalho, de que o padrão mais recente, H.265/HEVC, detém maior eficiência na compressão de vídeos. Busca-se comprovar esta expectativa por meio do emprego de métricas de avaliação da codificação de vídeos, originalmente em formato YUV.

Selecionando-se dois codificadores de cada um dos padrões, além de softwares para reprodução de cada um dos tipos de codificação, foi proposto fazer uma análise empregando as métricas PSNR e SSIM.

Durante o processo de codificação/decodificação dos vídeos, encontrou-se a dificuldade em se aplicar as métricas previamente mencionadas, pelo fato de não ter sido encontrado uma forma que permitisse medir os valores por meio dos softwares utilizados. No entanto, foram apresentadas as taxas de compressão de cada codificador, evidenciando maiores taxas para o padrão H.265/HEVC.

## 5.2 QUANTITATIVE PERFORMANCE EVALUATION OF THE EMERGING HEVC/H.265 VIDEO CODEC

No artigo de Koumaras et al (2012), faz-se uma comparação quantitativa de desempenho do padrão HEVC. O objetivo é pesquisar se a nova codificação cumpre com seu objetivo principal, de dobrar a compressão sem degradação da qualidade da codificação. Para tanto, utiliza-se o padrão anterior AVC como parâmetro de comparação.

Introduz-se apresentando as expectativas sobre o novo padrão HEVC, assim como, descrevendo-se a abordagem de blocos utilizada e seus diversos níveis de abstração. Descreve-se, também, em termos gerais, a predição *Intra* e *Inter*, modificações na codificação por entropia, perfis e a própria métrica quantitativa sobre qualidade PSNR.

Utiliza-se, no processo de comparação, diversos parâmetros como taxa de bits, passos de quantização e eficiência de compressão. Utiliza-se de tabelas e gráficos que evidenciam o melhor desempenho do HEVC, principalmente para passos de quantização maiores. Apresenta-se, também, uma comparação entre quadros dos dois padrões que demonstram uma diferença abrupta de qualidade para um passo de quantização elevado.

Conclui-se que o padrão HEVC cumpre com seus objetivos e, realmente se mostra mais eficiente.

## 5.3 PERFORMANCE COMPARISON OF H.265/MPEG-HEVC, VP9, AND H.264/MPEG-AVC ENCODERS

No artigo de Grois et al (2013), uma comparação de desempenho entre *codecs* dos padrões H.265/MPEG-HEVC, VP9, e H.264/MPEG-AVC é empregada. Teve-se como objetivo obter informações da comparação de desempenho das *codecs* dos padrões da ITU-T/ISO-EIC em relação à *codec* do padrão VP9, desenvolvido pela Google.

As *codecs* utilizadas foram o modelo de referência da ITU-T, HM Reference Software Encoder, a biblioteca x.264 e a *codec* VP9, para os respectivos padrões H.265/MPEG-HEVC, H.264/MPEG-AVC e VP9.

Manteve-se parâmetros de execução equivalentes para cada codificador, onde tabelas são apresentadas para os arquivos de configuração para cada um deles.

Comparou-se as implementações baseando-se em métricas, tais como: taxa de bits, PSNR e complexidade computacional.

Chegou-se a conclusão de que o codificador do padrão H.265/MPEG-HEVC é superior tanto em comparação com seu antecessor H.264/AVC, quanto em comparação do da Google, o padrão VP9.

#### 5.4 MPEG VIDEO DEPLOYMENT IN INTERACTIVE MULTIMEDIA SYSTEMS: HEVC VS. AVC CODEC PERFORMANCE STUDY

No artigo apresentado em 2013, no congresso WSEAS Transactions on Signal Processing, os autores Milovanovic e Bojkovic (2013), da Universidade de Belgrado, Sérvia, conduziram uma comparação de desempenho entre codificadores dos padrões H.265/HEVC, H.264/AVC e H.263.

É apontado inicialmente as diferenças entre estas tecnologias de codificação desenvolvidas pela ITU-T. Cada padrão é comparado em termos de avanço tecnológico no que diz respeito a suas implementações e técnicas para compressão dos vídeos. É evidenciado ao longo do texto, as ferramentas de codificação mais sofisticadas do padrão H.265/HEVC.

O artigo também aborda sobre aplicações que exigem baixa latência dos codificadores, como vídeo conferência. Da mesma forma, as implementações dos padrões utilizadas para efetuar a comparação são discutidas e apresentadas.

As métricas utilizadas durante o processo de análise, foram as contabilizações de redução de taxa de bits, eficácia da codificação e complexidade computacional. Os valores obtidos comprovaram a superioridade do padrão H.265/HEVC que se saiu consideravelmente melhor nos testes.

## 5.5 AVALIAÇÃO DE DESEMPENHO DA ESCALABILIDADE ESPACIAL NA RECOMENDAÇÃO H.264/SVC PARA CODIFICAÇÃO DE VÍDEO

Este projeto, a monografia do então graduando no curso de Bacharel em Ciência da Computação da UNESC, Márton Henrique de Souza, discute o desempenho do H.264/SVC, em especial a sintonia de quantização para melhor explorar as condições de escala espacial. A quantização de vídeos digitais varia de 0 a 51, as simulações realizadas visaram avaliar a influência da quantização neste intervalo, de acordo com métricas como taxa de bits, PSNR e complexidade computacional. Os resultados comprovaram que as condições de quantização da camada base de um vídeo com escalabilidade espacial, influenciam diretamente na taxa de bits e no tempo de codificação do vídeo resultante.

## 5.6 SEGMENTAÇÃO QUADTREE PARA SELEÇÃO DE BLOCOS NA CODIFICAÇÃO DE VÍDEO H.264/AVC

Este projeto, a monografia do então graduando no curso de Bacharel em Ciência da Computação da UNESC, Tiago Pizzetti Medeiros, discute a utilização de uma estrutura de segmentação baseada em quadtree para seleção dos macroblocos a serem codificados pelo padrão H.264/AVC. Nessa estrutura, os macroblocos das imagens de diferença para a estimação e a compensação de movimento são segmentados em estruturas regulares de 16x16, 8x8 e 4x4 pixels. Os resultados indicam que o esquema de codificação descrito apresenta complexidade computacional reduzida em relação à estrutura do Padrão H.264/AVC convencional, mantendo-se equivalente a relação taxa-qualidade.

## 5.7 DISTRIBUIÇÃO DE VÍDEO ESCALÁVEL BASEADA NA RECOMENDAÇÃO H.264/SVC EM UM SERVIÇO DE WEBTV

Desenvolveu-se nesta monografia do então graduando no curso de Bacharel em Ciência da Computação da UNESC, Luiz Felipe Lazzarin, um protótipo de codificação escalável espacialmente baseado no software de referência Joint Scalable Video Model e decodificado/apresentado pela biblioteca OpenSVCEncoder, ambos compatíveis com o padrão H.264/SVC, que demonstra a eficiência do serviço. Constatou-se a indisponibilidade de soluções modulares - com

licenças de uso livres ou proprietárias – de *codecs* para que se possa desenvolver uma solução própria dentro do serviço de WebTV. Implementou-se um serviço WebTV baseado em HTML5, que serve de repositório para vídeos e também instancia streamings ao vivo. A referida implementação visa oferecer um serviço de qualidade em disponibilização de programação de vídeo em escala e em transmissões ao vivo.

## 6 ANÁLISE COMPARATIVA ENTRE OS PADRÕES ITU-T H.265/HEVC E H.264/AVC EM CODIFICAÇÃO DE VÍDEO

O presente trabalho propôs-se a efetuar uma análise comparativa de desempenho e complexidade computacional do tipo objetiva-quantitativa entre os dois padrões mais recentes em codificação de vídeo desenvolvidos pelo setor de padronização da *International Telecommunication Union* (ITU-T), e pelo ISO/IEC *Moving Pictures Experts Group* (MPEG). São conhecidos pelas siglas H.264, ou *Advanced Video Coding* (AVC), e H.265, ou *High Efficient Video Coding* (HEVC).

Estes padrões propõem a padronização da sintaxe presente no fluxo de dados interno de uma *codec* de vídeo. Desta forma, procura-se fornecer uma regulamentação para que todos os vídeos codificados por um codificador – sendo este implementado de acordo com o padrão -, possam ser corretamente interpretados e decodificados. No entanto, os padrões, por si só, não garantem a qualidade de reprodução *end-to-end*.

Fazem parte da arquitetura destes padrões, diferentes perfis e respectivos níveis, os quais parametrizam o processo da codificação. Através desta parametrização, o resultado final da codificação pode ser adequado com demandas específicas de vídeo, variando em termos de latência e qualidade final de reprodução. Assim, ambas as tecnologias são capazes de atender com eficiência e flexibilidade as expectativas de vídeos de alta qualidade.

O padrão H.264/AVC foi apresentado em 2003, tendo-se demonstrado efetivo até os dias atuais. Ele é utilizado em diversos tipos de serviços de vídeo, desde televisão digital, cinema digital, vídeo teleconferência, vídeos da Internet, serviços de compartilhamento de vídeo – como Youtube, por exemplo - , Vídeo sob Demanda (*Video on Demand* - VoD), vídeos caseiros/amaradores e discos Blu-ray. No entanto, é perceptível que a demanda por vídeos de resolução cada vez maior cria a necessidade do desenvolvimento de tecnologias que cumpram com tais expectativas. Em 2013, o padrão H.265/HEVC foi apresentado ao mundo. Desde então, diversos estudos têm sido realizados para avaliar seu desempenho, muitas vezes comparativamente ao seu antecessor H.264/AVC, ou mesmo à outros codificadores – como o VP9, da Google.

Grande número de análises em codificação de vídeo são efetuadas em termos de desempenho e complexidade computacional. Ambos os parâmetros possuem duas formas de serem avaliados, objetivamente ou subjetivamente. Para

tanto, em ambos os casos, podem ser empregadas métricas quantitativas ou qualitativas, por meio das quais são efetuadas medições durante e após o processo de codificação/decodificação. A partir deste ponto é possível comparar os resultados com o arquivo original não comprimido. O sub-item seguinte, aprofunda estes conceitos e apresenta as métricas objetivas empregadas no presente trabalho.

No mais, este capítulo procurará apresentar a forma como a presente pesquisa foi conduzida, os parâmetros e configurações adotados, metodologia empregada, formas de coleta de dados, apresentação de resultados, e por fim, discussão e conclusões finais.

## 6.1 METODOLOGIAS EM ANÁLISES DE CODIFICAÇÃO DE VÍDEO

Neste subcapítulo serão descritas as abordagens existentes para pesquisas em codificação de vídeo (subitens 6.1.1 e 6.1.2), assim como a metodologia efetivamente empregada na análise comparativa desenvolvida no presente trabalho (6.1.3).

### 6.1.1 Abordagem Objetiva

Uma análise objetiva em codificação de vídeo se refere ao uso de modelos matemáticos (métricas) para mensurar diferentes aspectos relacionados ao desempenho do processo de codificação/decodificação, e, a qualidade final do arquivo codificado, baseando-se no ruído presente no sinal. A seguir são descritas as principais métricas quantitativas, as quais foram utilizadas no presente trabalho.

#### 6.1.1.1 Taxa de bits

A *taxa de bits* refere-se ao fluxo de dados necessário para reproduzir um vídeo digital em determinada resolução. Quanto maior for a taxa de bits necessária, maior custo computacional - no sentido de, processamento, armazenamento e transmissão deste sinal. Portanto, um dos principais objetivos dos codificadores de vídeo é, justamente, diminuir a taxa de bits para reprodução de vídeos (SULLIVAN; WIEGAND, 2005, tradução).

A qualidade de um vídeo, assim como a eficiência de um codificador, será tão quanto maior para *taxas de bits* cada vez menores produzidas por sua saída,

mantendo-se a fidelidade ao vídeo original. Sendo assim, esta é uma métrica essencial na avaliação objetiva de *codecs* de vídeo e, é comumente medida em kilobits por segundo (*kbits/s*).

É válido dizer que, relacionado à *taxa de bits* está a taxa de compressão de um arquivo de vídeo. Podemos obter o tamanho do arquivo final comprimido através da fatoração entre *taxa de bits* e *tempo de execução* do codificador. Este resultado nos permite comparar com o tamanho de armazenamento do arquivo original (não-comprimido), donde é possível estabelecer um percentual conhecido como *taxa de compressão*.

#### 6.1.1.2 Complexidade Computacional

Existem diferentes formas de se medir a complexidade computacional de um codificador de vídeo. O parâmetro *tempo de processamento*, tem sido amplamente utilizado para a medição. Trata-se de uma cronometragem do processo de codificação como um todo. Geralmente, os modelos de referência geram relatórios ao término contendo o *tempo de processamento*.

Porém, outras abordagens como análise teórica, análise direta de código de referência e avaliação de consumo de energia, também têm sido aplicadas para medir complexidade (CORREA et al, 2012, tradução nossa).

Não existe um consenso por qual método utilizar para efetuar tais medições. É sabido que estes diferentes métodos de medição estão passíveis de falta de acurácia para as medições.

#### 6.1.1.3 Razão Sinal-Ruído de Pico (PSNR)

A *Razão de Ruído para Pico de Sinal* (PSNR), é uma métrica quantitativa que mede a qualidade, amplamente utilizada em análises de codificação de vídeo. Trata-se da medida da quantidade de ruído presente no sinal de um vídeo digital. Seu valor é medido em *decibéis* (*dB*), e pode ser obtido através da seguinte fórmula matemática:

$$\text{PSNR} = 10 \log_{10}(255^2 / \text{MSE})$$

onde MSE é o erro quadrático médio entre os valores das amostras originais e decodificadas correspondentes (SULLIVAN; WIEGAND, 2005, tradução). Tem-se, portanto, a seguinte fórmula:

$$MSE = \frac{1}{M \cdot N} \sum_{i=0}^M \sum_{j=0}^N (f_{i,j} - \hat{f}_{ij})^2$$

Onde M é o número de Linhas; N é o número de colunas.  $f_{i,j}$  é o valor de cada pixel de um quadro original e  $\hat{f}_{ij}$  é o valor de cada pixel de um quadro reconstruído (codificado e decodificado).

### 6.1.2 Abordagem Subjetiva

Existem diferentes métricas utilizadas para efetuar uma análise subjetiva de qualidade em codificação de vídeo. Muitas delas, no entanto, embora proponham uma abordagem de avaliação mais próxima ao funcionamento e sensibilidade do sistema visual humano, encontram problemas em termos de confiabilidade.

O único método confiável para avaliar a qualidade de um vídeo observado por um humano, é perguntar a pessoas pela sua opinião, o que é chamado de Avaliação Subjetiva da Qualidade de Vídeo (VQA). VQA-subjetiva é impraticável para a maioria das aplicações devido ao envolvimento humano no processo. No entanto, estudos VQA-subjetivos fornecem dados importantes para avaliar o desempenho de métodos objetivos ou automatizados de avaliação de qualidade. Além de prover os meios para avaliar o desempenho do estado-de-arte de tecnologias VQA, estudos subjetivos também permitem melhorias no desempenho de algoritmos VQA, no sentido de alcançar a meta final de equiparação com a percepção humana (KALPANA et al, 2009, tradução nossa).

### 6.1.3 Metodologia empregada no presente trabalho

A abordagem utilizada neste trabalho foi do tipo objetiva, fazendo-se uso das métricas quantitativas anteriormente descritas. A seguir, encontra-se as sequências de vídeo utilizadas para os testes e como se deu a coleta de dados.

### 6.1.3.1 Seleção e características das sequências de vídeo utilizadas para teste

As sequências utilizadas nos testes foram obtidas em um *website* pertencente a Universidade do Estado do Arizona (USA), que disponibilizam-nas livremente para fins acadêmicos. O endereço pode ser encontrado nas referências desta monografia.

As sequências caracterizam-se por estarem em formato original, ou seja, um formato planar sem compressão, no caso, YUV.

Foram utilizadas sequências com resoluções de 176 x 144 *pixels*, também conhecido como formato “qcif”. A tabela 3, lista os arquivos utilizados, sua *resolução*, *sub-amostragem de cores e extensão*.

Tabela 1 – Lista de sequências utilizadas nos testes e suas características.

Nome da sequência	Resolução	Sub-amostragem	Classificação	Formato
foreman	176x144	4:2:0	qcif	YUV
miss-america	176x144	4:2:0	qcif	YUV
salesman	176x144	4:2:0	qcif	YUV

Fonte: Autor.

A sequência *Foreman* caracteriza-se por ter um alto grau de movimentação tanto de fundo como de frente ao longo de seus quadros. Esta é uma sequência que exige do codificador considerável quantidade de processamento destinado a predição *Inter*.

Figura 19 – Sequência Foreman em formato YUV



Fonte: Autor.

A sequência *Miss-américa* possui um fundo preto e fixo, tendo à frente pequenas mudanças de movimentação referentes à pessoa que está falando. Fazendo-se uso desta sequência, é possível testar nos codificadores a capacidade de tirar vantagem de uma área relativamente grande, a qual é estática e de pouca variação cromática.

Figura 20 – Sequência Miss-américa em formato YUV



Fonte: Autor.

A sequência *Salesman*, possui um fundo fixo com movimentação de frente e sombra, além de considerável quantidade de detalhes na cena. Tal sequência permite testar o nível de granularidade resultante do processo de codificação codificadores em relação as diferentes cores e texturas presentes.

Figura 21 – Sequência Salesman em formato YUV



Fonte: Autor.

Deve-se dizer que, as sequências listadas são amplamente utilizadas em pesquisas em codificação de vídeo, o que reafirma a relevância da pesquisa aqui efetuada e a possibilidade de comparação dos resultados aqui obtidos com pesquisas em codificação de vídeo ao redor do mundo.

Para visualização dos vídeos originais e reconstruídos, foi utilizado o software “YUVplayer” (vide referências), o qual é livre para uso acadêmico. Este software

permite reproduzir vídeos em formatos planares (não-codificados) e ainda especificar certas configurações para execução, como *resolução*, *taxa de quadros* e *formato de sub-amostragem* das cores.

Para visualização dos arquivos de vídeos codificados foi utilizado o software de código aberto VLC Media Player (vide referências), o qual é livre.

### 6.1.3.2 Coleta de dados

A coleta de dados foi razoavelmente simples de se fazer para ambos os modelos de referência. Os softwares por si só forneceram como saída as informações necessárias para se efetuar as medições de desempenho e complexidade computacional.

Portanto, a saída do codificador de ambos os padrões foi redirecionada da saída padrão (tela) para arquivos de texto que foram salvos no disco rígido. A partir daí, utilizou-se de shell scripts para manipular os arquivos e extrair linhas inteiras que continham os dados relevantes. Uma segunda etapa foi aplicada na extração dos dados, fazendo-se uso de *scripts* escritos na linguagem *Perl*, a qual foi projetada para processamento de texto e possui sintaxe bastante simples no tratamento de arquivos de texto.

Para posterior geração de gráficos de comparação de desempenho e complexidade computacional, foi utilizado o programa por linha de comando *gnuplot*, o qual é livre para uso.

## 6.2 CENÁRIO DA ANÁLISE COMPARATIVA EFETUADA

### 6.2.1 Implementações utilizadas

Buscou-se na Internet implementações dos padrões em análise e foram encontradas implementações da própria ITU-T, as quais foram desenvolvidas como modelos de referência – JM Reference Model (H.264/AVC) e HM Reference Model (H.265/HEVC) -. No entanto, estas implementações procuram empregar todas as técnicas de codificação de vídeo elucidadas nos capítulos anteriores, de acordo com a norma de cada padrão. O endereço em que tais foram encontradas estão incluídos nas referências deste trabalho.

Sendo estas implementações desenvolvidas pelo próprio órgão padronizador e livres de interesse comercial ou, de customização de terceiros, acreditou-se serem as melhores opções de softwares existentes para se efetuar uma análise imparcial dos padrões.

### **6.2.2 Ambiente de execução e configurações iniciais de software**

Foi utilizado o sistema operacional Linux, distribuição Ubuntu 14.04, versão 64 bits. O terminal *Bash* do sistema operacional foi utilizado para efetuar todos os testes, tanto de codificação como de decodificação, tanto para o *JM Reference Model* como para o *HM Reference Model*.

O sistema operacional possui nativamente, bibliotecas das linguagens C e C++, as quais foram utilizadas para a compilação dos softwares. A partir deste momento em diante, foi possível especificar arquivos de entrada e demais parâmetros de execução com arquivos de configuração, para, então, realizar os testes. A compilação dos softwares geram executáveis separados para codificação e decodificação.

Foi utilizado um laptop pessoal com as seguintes configurações de hardware: Intel® Core™ i7-4500U 1.80GHz de quatro núcleos e 8 GB de memória RAM.

Ambos os modelos de referência possuem documentação clara que instrui desde a instalação até execução dos softwares.

### **6.2.3 Parâmetros de execução e arquivos de configuração**

A sintaxe dos comandos e os parâmetros de execução, são, também, detalhados na documentação que vem juntamente com os códigos-fonte.

Arquivos de configuração padrão – tanto para o codificador como para o decodificador - também são fornecidos para cada software. Dentro destes arquivos estão especificados todos os parâmetros de execução em forma de texto. Foram mantidos a maioria dos valores já existentes, fazendo-se as alterações necessárias para cada arquivo de entrada e *taxa de bits alvo*.

Abaixo, encontram-se a tabela 1 e tabela 2, as quais apresentam os parâmetros de execução que foram alterados – tanto para o *JM Reference Model*

como para o *HM Reference Model*, respectivamente – para a sequência de vídeo “foreman” (vide item 6.3.1).

Tabela 2 – Parâmetros de execução fornecidos ao codificador do software *JM Reference Model*.

Parâmetro	Valores
InputFile	= tcc2/sequences/yuv/foreman_qcif.yuv
SourceBitDepthLuma	= 8 # Source Bit Depth for Luma color component (8...14 bits)
SourceBitDepthChroma	= 8 # Source Bit Depth for Chroma color components (8...14 bits)
FrameRate	= 30 # Frame Rate per second
FrameSkip	= 0 # Number of frames to be skipped in input
SourceWidth	= 176 # Input frame width
SourceHeight	= 144 # Input frame height
FramesToBeEncoded	= 300 # Number of frames to be coded
IntraPeriod	= 32 # Period of I-pictures (0=only first)
RateControlEnable	= 1 # 0 Disable, 1 Enable
Bitrate	= 32000 # Bitrate(bps)
RCUpdateMode	= 2 # Rate Control type.
TraceFile	= "tcc2/trace_enc.txt" # Trace file
ReconFile	= "tcc2/test_rec.yuv" # Reconstruction YUV file
OutputFile	= "tcc2/test.264" # Bitstream
StatsFile	= "tcc2/stats.dat" # Coding statistics file

Fonte: Autor.

Tabela 3 – Parâmetros de execução fornecidos ao codificador do software *HM Reference Model*.

Parâmetro	Valores
InputFile	: tcc2/sequences/yuv/foreman_qcif.yuv
FrameRate	: 30 # Frame Rate per second
FrameSkip	: 0 # Number of frames to be skipped in input
SourceWidth	: 176 # Input frame width
SourceHeight	: 144 # Input frame height
FramesToBeEncoded	: 300 # Number of frames to be coded
Level	: 3.1
RateControl	: 1 # Rate control: enable rate control
TargetBitrate	: 32000 # Rate control: target bitrate, in bps
BitstreamFile	: tcc2/str.bin
ReconFile	: tcc2/rec.yuv

Fonte: Autor.

Foi definido para todos os testes:

- a) *profundidade em bits* para cada pixel de cor no valor de 8 bits;
- b) *taxa de quadros* no valor de 30 quadros por segundo;
- c) *período de codificação Intra* a cada 32 quadros.

Já para os *perfis* e *níveis* de cada um dos softwares, foi utilizada a seguinte configuração:

a) *JM Reference Model* – Perfil: *Main* ; nível: 4.0;

b) *HM Reference Model* – Perfil: *Random Access Main* ; nível: 3.1.

Os parâmetros de execução apresentados anteriormente (tabelas 1 e 2) referem-se ao teste efetuado com a *taxa de bits alvo* no valor de 32 kbps, o equivalente a 32.000 bps, como especificado. Abaixo, na figura 19 lista-se todos os testes realizados.

Figura 22 – Testes e taxas de bits alvo utilizadas.

Teste	Taxa de bits alvo					
	1	2	3	4	5	6
<b>JM Reference Model</b>	32 kbps	64 kbps	128 kbps	256 kbps	512 kbps	1024 kbps
<b>HM Reference Model</b>	32 kbps	64 kbps	128 kbps	256 kbps	512 kbps	1024 kbps

Fonte: Autor.

### 6.3 DESAFIOS ENCONTRADOS

Os desafios encontrados durante a condução da pesquisa do presente trabalho se deram em todas as suas fases, desde o levantamento bibliográfico até a geração dos gráficos de comparação de desempenho e complexidade. Isso se deve ao conteúdo estudado, o qual é denso e bastante desafiador, pois envolve diferentes conceitos associados a múltiplos campos do conhecimento.

Sendo assim, foi-se dedicado um tempo considerável para o levantamento bibliográfico, o qual possuía predominância de artigos científicos escritos na língua inglesa. Durante a prospecção, foram encontrados uma quantidade considerável de trabalhos relevantes que necessitavam ser pagos para serem acessados, embora todos os artigos referenciados no presente trabalho são encontrados com livre acesso.

Tendo-se apropriado dos conceitos e dos próprios padrões de codificação de vídeo, fez-se uma busca extensiva por implementações dos padrões em análise, assim como sequências de vídeo para teste e softwares capazes de reproduzi-las. A definição de como os testes seriam conduzidos foi algo que necessitou planejamento e embasamento teórico.

Deve-se por algum tempo na compilação dos softwares e execução dos mesmos. A princípio, foram compilados no sistema operacional Fedora, mas apresentaram problemas de execução graves. Viu-se que poderia ser vantajoso mudar o ambiente de execução, alternando para o sistema operacional Ubuntu, no qual foi possível compilar e executar os softwares sem maiores complicações.

O processo de codificação e decodificação das sequências caracterizou-se por ser um trabalho árduo, uma vez que foram feitos diversos testes com várias sequências de vídeo, incluindo, obviamente, as selecionadas para análise neste trabalho.

Logo a seguir, o processo de extração de dados e coleta de informação requereu conhecimentos adicionais do acadêmico para lidar com os arquivos de texto e a geração dos gráficos.

Por fim, o processo de análise dos resultados e término da escrita deste monografia necessitou de constante apoio do orientador e planejamento.

## 6.4 RESULTADOS

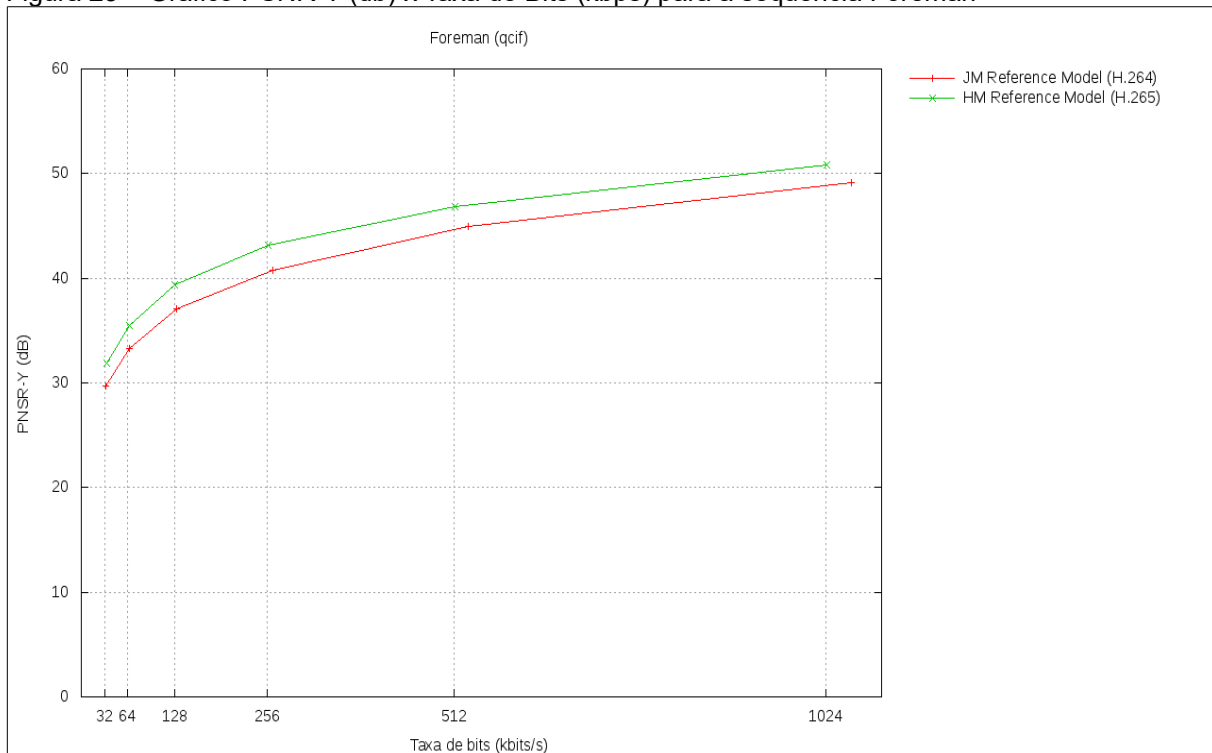
Os resultados obtidos do processo da codificação das sequências anteriormente mencionadas são apresentados ao longo deste subcapítulo. Foram gerados dois tipos de gráficos para efetuar a comparação de desempenho e complexidade computacional entre *codecs* analisadas.

O gráfico de *PSNR-Y* em função da *taxa de bits*, relaciona o aumento da qualidade de um vídeo baseado na largura do canal de dados, o qual proporciona melhores resultados quanto mais elevado for. Portanto, a *taxa de bits* influencia diretamente na *taxa de compressão* do arquivo de vídeo original. Taxas de bits menores forçam as *codecs* a adaptarem o fluxo de dados de uma sequência de vídeo, descartando mais informações relativamente a taxas-alvo de bits maiores.

O gráfico de *complexidade computacional* em função da *taxa de bits*, relaciona o custo computacional durante o processo de codificação em termos de unidades de tempo – no caso, segundos (s) – baseado, também, na largura do canal de dados. A *complexidade computacional* tende a se elevar proporcionalmente ao aumento das taxas-alvo de bits, devido ao aumento de dados disponíveis a serem manipulados – aqui refere-se a aplicação de técnicas de compressão de sinal digital: quantização, entropia e transformadas - e preditos.

Na figura 23, é possível observar o desempenho de ambos os modelos de referência para a sequência *Foreman*.

Figura 23 – Gráfico PSNR-Y (db) x Taxa de Bits (kbps) para a sequência Foreman



Fonte: Autor.

Ambos os modelos de referência são representados no gráfico em escala logarítmica. É possível ver que o modelo de referência do padrão H.265/HEVC é superior em desempenho e qualidade.

Abaixo, na figura 24, apresentam-se alguns quadros da sequência decodificados, para a taxa mínima definida para os testes de 32 kbps e a máxima, definida em 1024 kbps.

Figura 24 – Quadros decodificados da sequência Foreman: (a) H.264/AVC 32 kbps (b) H.264/AVC 1024 kbps (c) H.265/HEVC 32 kbps (d) H.265/HEVC 1024 kbps.



Fonte: Autor.

Na figura 25, o mesmo quadro, original, como referência.

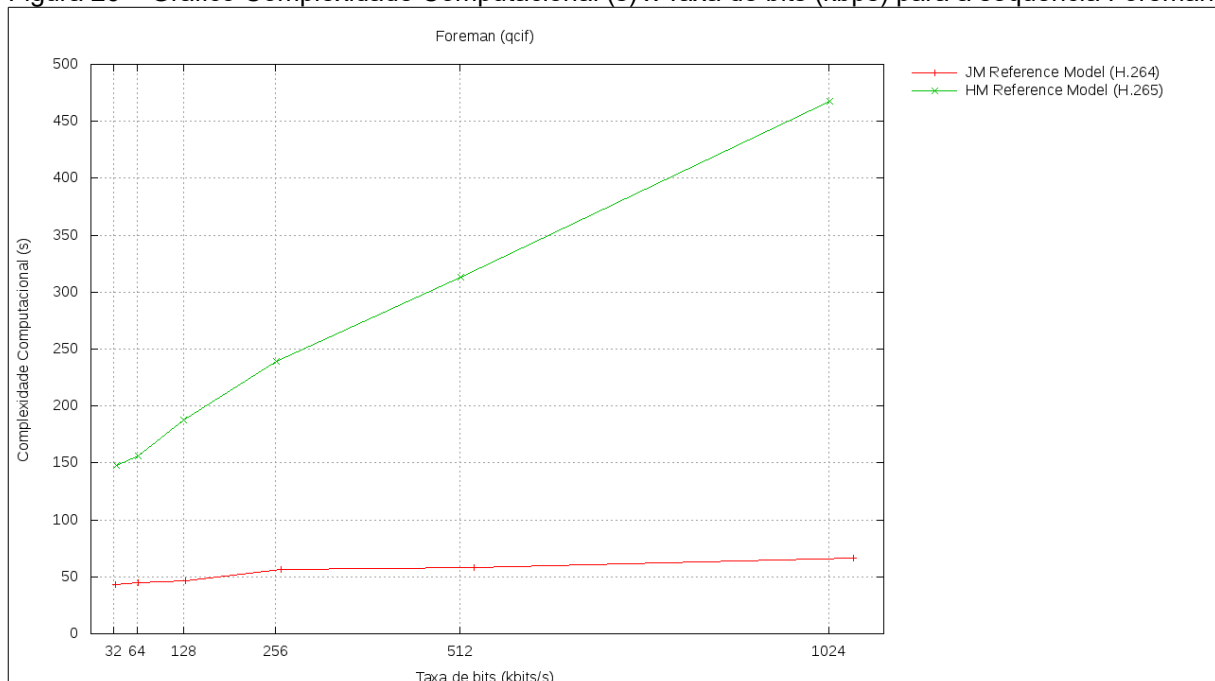
Figura 25 – Quadro original para a sequência Foreman.



Fonte: Autor.

Já na figura 26, é possível observar o custo computacional de ambos os modelos de referência para a sequência *Foreman*.

Figura 26 – Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Foreman

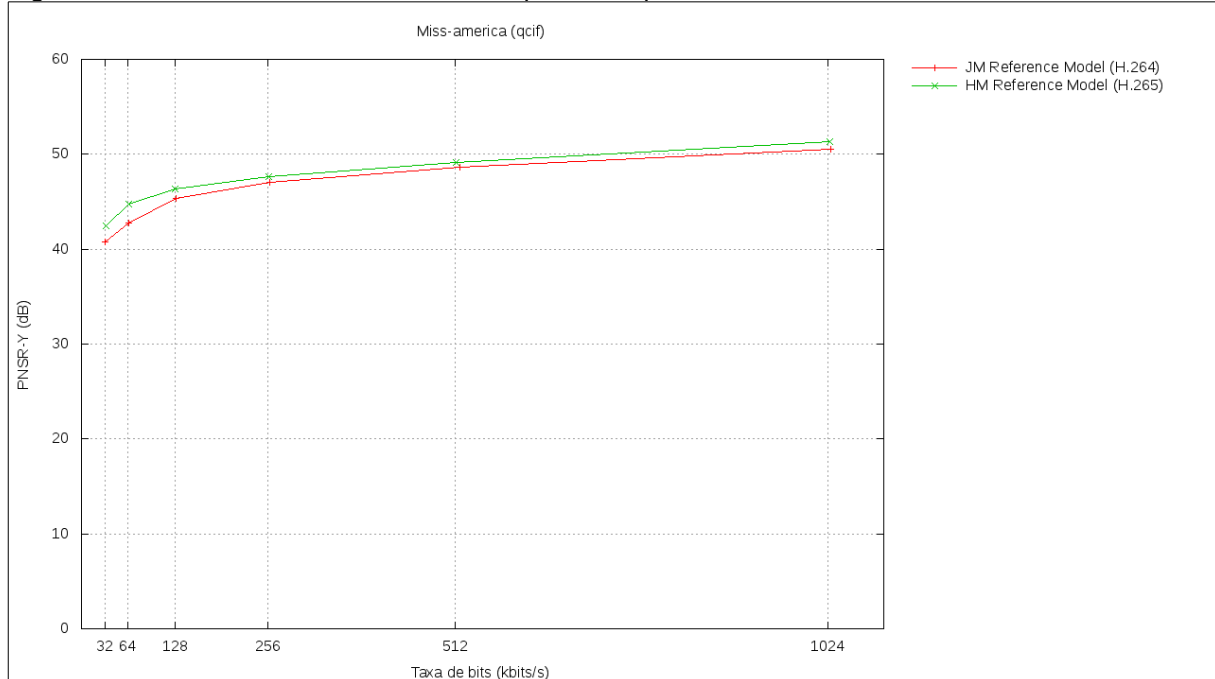


Fonte: Autor.

Percebe-se que a complexidade computacional do modelo de referência do padrão H.265/HEVC foi consideravelmente maior, apresentando-se com um crescimento constante retilíneo. Por outro lado, o modelo de referência do padrão H.264/AVC, apresentou, relativamente, valores e crescimento consideravelmente menores.

Na figura 27, é possível observar o desempenho de ambos os modelos de referência para a sequência *Miss-américa*.

Figura 27 – Gráfico PSNR-Y x Taxa de Bits para a sequência Miss-américa



Fonte: Autor.

Ambos os modelos de referência são representados no gráfico em escala logarítmica. É possível ver que o modelo de referência do padrão H.265/HEVC é, novamente, superior em desempenho e qualidade, embora que, para esta sequência em específico, os modelos alcançam valores bem próximos. Acredita-se que tal resultado deve-se a grandes porções estáticas e de mesma coloração nos quadros ao longo sequência.

Abaixo, na figura 28, apresentam-se alguns quadros da sequência decodificados, para a taxa mínima definida para os testes de 32 kbps e a máxima, definida em 1024 kbps.

Figura 28 - Quadros decodificados da sequência Miss-américa: (a) H.264/AVC 32 kbps (b) H.264/AVC 1024 kbps (c) H.265/HEVC 32 kbps (d) H.265/HEVC 1024 kbps.



Fonte: Autor.

Na figura 29, o mesmo quadro, original, como referência.

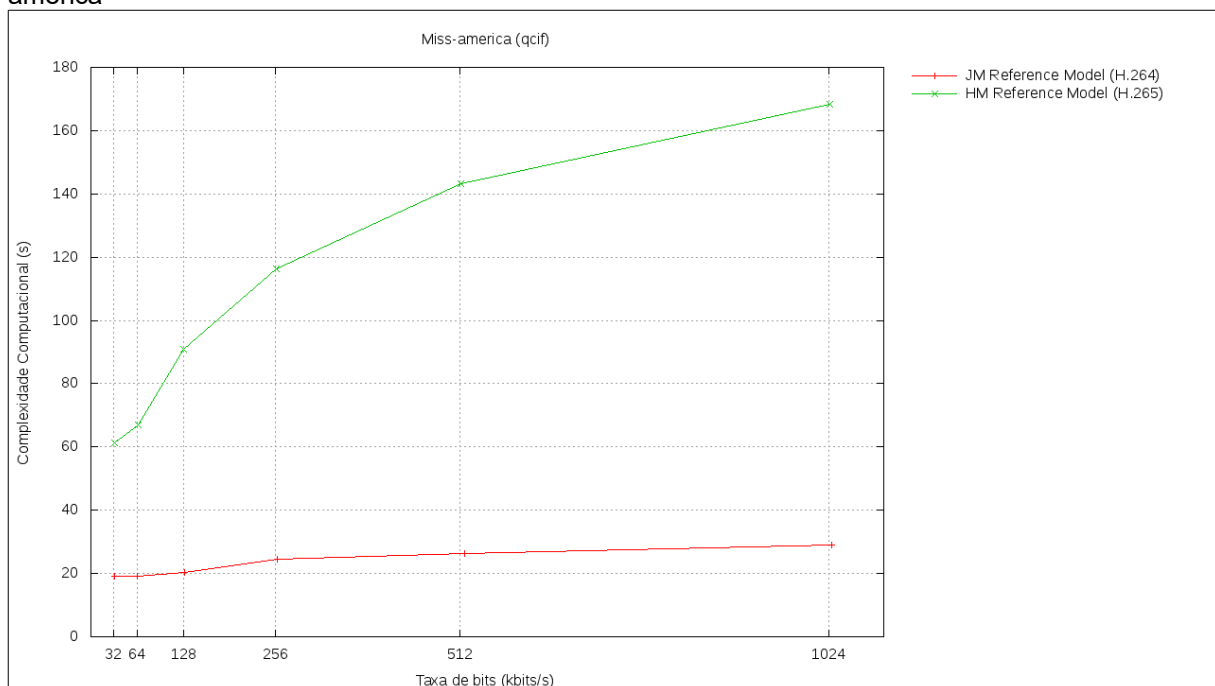
Figura 29 – Quadro original para a sequência Miss-américa.



Fonte: Autor.

Já na figura 30, é possível observar a complexidade computacional de ambos os modelos de referência para a sequência *Miss-américa*.

Figura 30 - Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Miss-américa



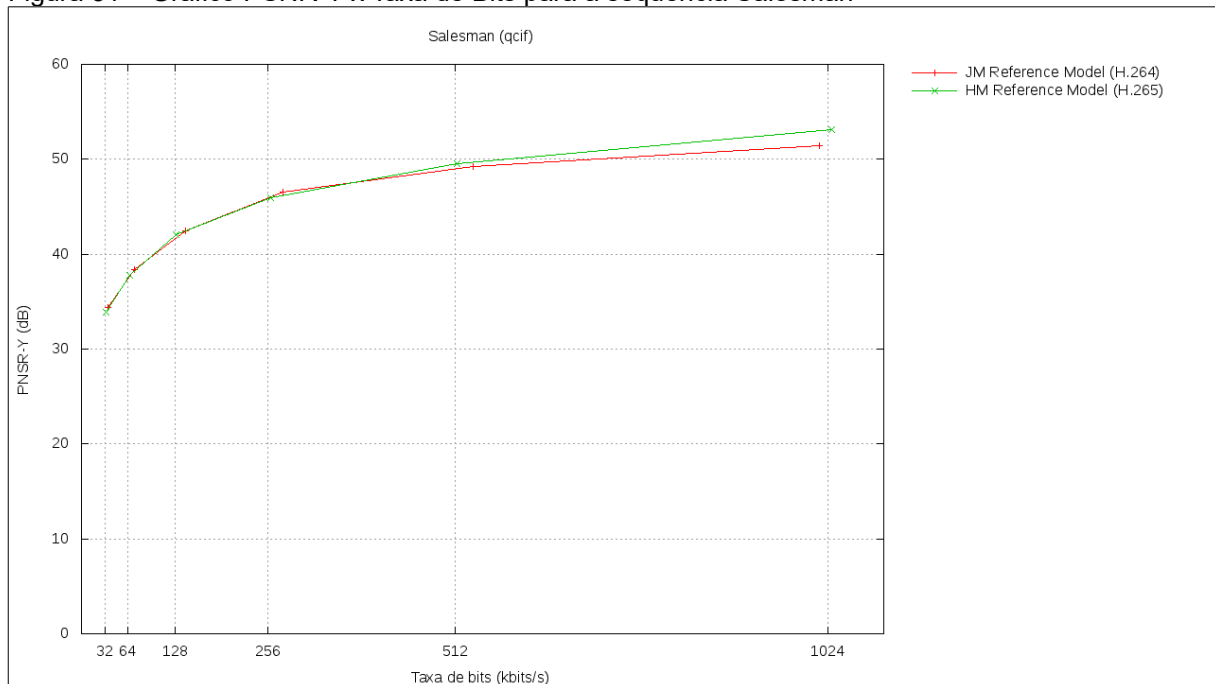
Fonte: Autor.

É possível ver que a complexidade computacional do modelo de referência do padrão H.265/HEVC foi consideravelmente maior, apresentando-se com um crescimento variante e decrescente. Por outro lado, o modelo de referência

do padrão H.264/AVC, novamente apresentou, relativamente, valores e crescimento consideravelmente menores.

Na figura 31, é possível observar o desempenho de ambos os modelos de referência para a sequência *Salesman*.

Figura 31 – Gráfico PSNR-Y x Taxa de Bits para a sequência *Salesman*



Fonte: Autor.

Ambos os modelos de referência são representados no gráfico em escala logarítmica. Desta vez, é possível ver que o modelo de referência do padrão H.265/HEVC é superior em desempenho e qualidade, apenas após a *taxa de bits* alcançar um valor mediano entre 256 kbps e 512 kbps. O modelo de referência do padrão H.264/AVC, praticamente equipara-se até este ponto em desempenho e qualidade, tendo uma leve desvantagem próximo a taxa de 256 kbps. No entanto, mais uma vez, o padrão H.265/HEVC se mostra superior para *taxas de bits* maiores.

Abaixo, na figura 32, apresentam-se alguns quadros da sequência decodificados, para a taxa mínima definida para os testes de 32 kbps e a máxima, definida em 1024 kbps.

Figura 32 - Quadros decodificados da sequência Salesman: (a) H.264/AVC 32 kbps (b) H.264/AVC 1024 kbps (c) H.265/HEVC 32 kbps (d) H.265/HEVC 1024 kbps.



Fonte: Autor.

Na figura 33, o mesmo quadro, original, como referência.

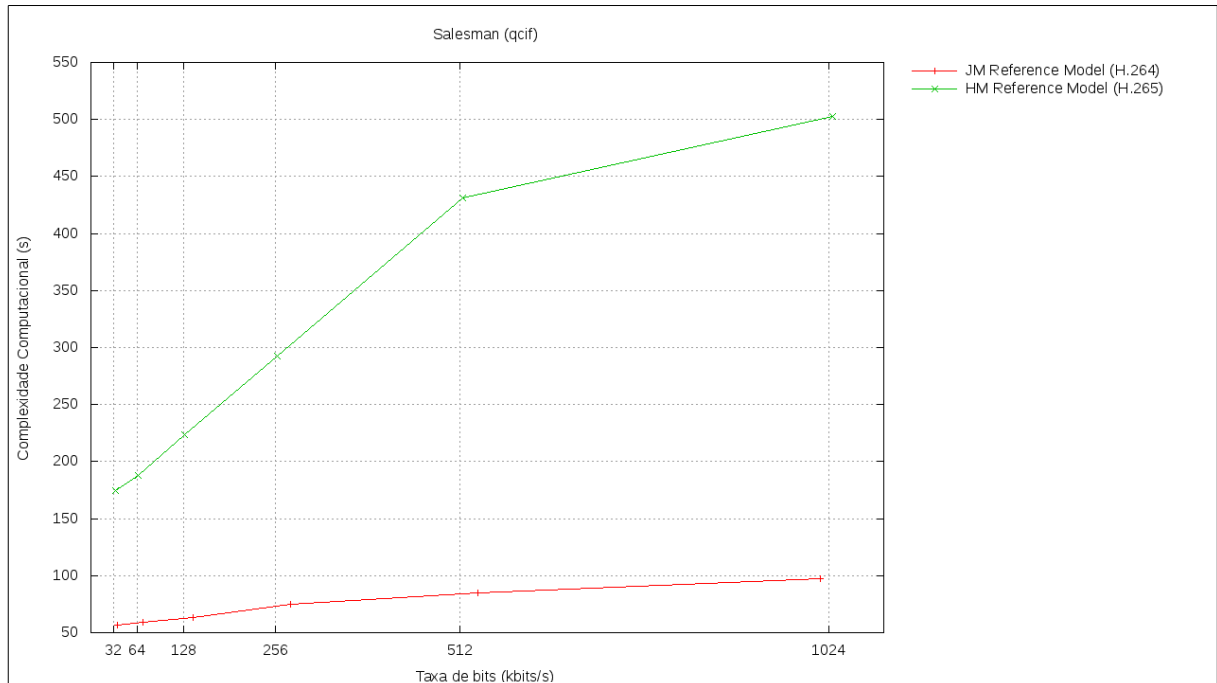
Figura 33 – Quadro original para a sequência Salesman.



Fonte: Autor.

Já na figura 34, é possível observar a complexidade computacional de ambos os modelos de referência para a sequência *Salesman*.

Figura 34 - Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência *Salesman*.



Fonte: Autor.

É visível que a complexidade computacional do modelo de referência do padrão H.265/HEVC foi, uma vez mais, consideravelmente maior, apresentando-se com um crescimento variante e decrescente. Por outro lado, o modelo de referência do padrão H.264/AVC, novamente apresentou, relativamente, valores e crescimento consideravelmente menores.

## 6.5 ANÁLISE DOS RESULTADOS

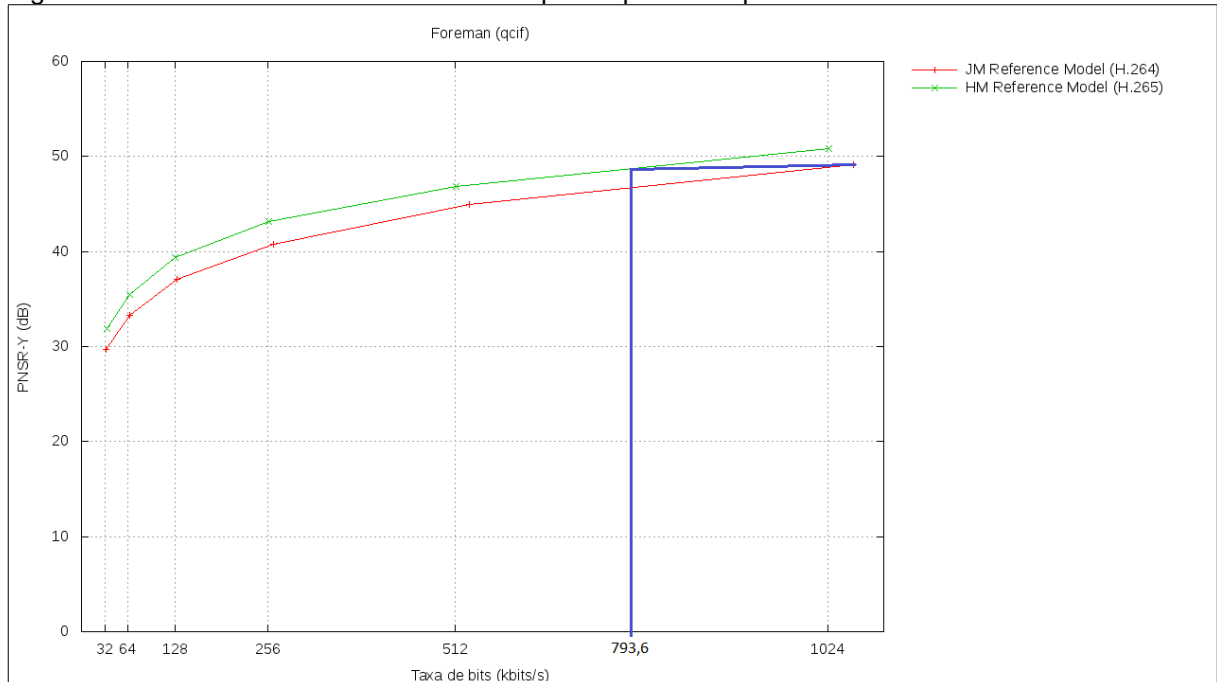
A partir dos gráficos de desempenho *PNSR-Y (db) x taxa de bits (kbps)* foi possível estabelecer pontos em cada gráfico nos quais a *codec* do padrão H.265/HEVC desponta em relação ao padrão H.264/AVC. Ou seja, pontos para os quais a codificação H.265/HEVC se torna mais interessante de ser empregada – no que se refere a desempenho – visto que para o padrão H.264/AVC existe uma tendência da qualidade se tornar constante mesmo com a elevação da taxa de bits.

Em outras palavras, visto que os resultados obtidos demonstram ao longo dos testes uma pequena diferença em qualidade (PSNR-Y) entre as *codecs* de ambos os padrões, estabelece-se um ponto em cada gráfico que marca o distanciamento do desempenho registrado pela *codec* do H.265/HEVC em relação ao desempenho registrado pela *codec* do H.264/AVC.

Foi traçada uma reta a partir do ponto máximo de qualidade (PSNR-Y) registrado pela *codec* do H264/AVC referente a taxa de 1024kbps, remetendo a um ponto na função registrada pela *codec* do H.265/HEVC.

Portanto, a intersecção das duas retas azuis nos gráficos abaixo, definem o ponto em que a *codec* do H.265/HEVC tende a ter um maior aumento de qualidade em relação à *codec* do H.264/AVC, à medida que se aumenta a taxa de bits. A seguir, nas figuras 35, 36 e 37, apresenta-se a análise como descrito para cada uma das sequências de vídeo.

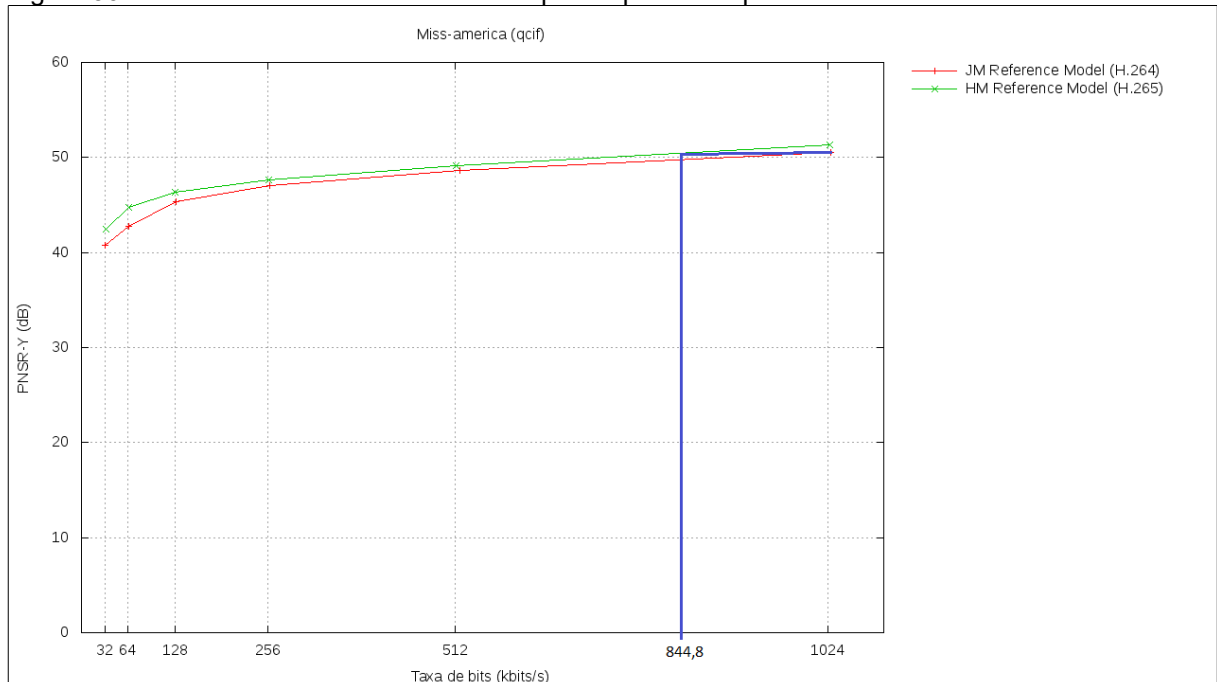
Figura 35 – Análise dos resultados de desempenho para a sequência Foreman.



Fonte: Autor.

Para a sequência *Foreman*, a projeção do ponto no eixo das abcissas equivale a *taxa de bits* no valor de 793,6 kbps.

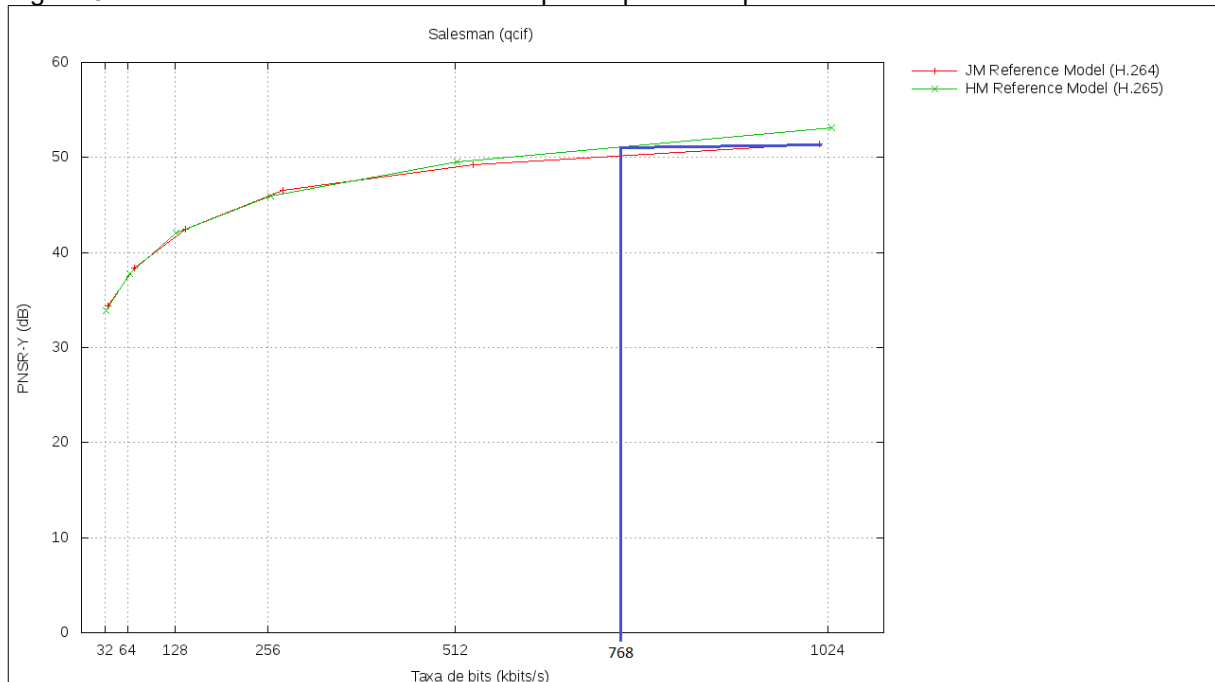
Figura 36 – Análise dos resultados de desempenho para a sequência Miss-américa.



Fonte: Autor.

Para a sequência *Miss-américa*, a projeção do ponto no eixo das abcissas equivale a *taxa de bits* no valor de 844,8 kbps.

Figura 37 – Análise dos resultados de desempenho para a sequência *Salesman*.



Fonte: Autor.

Para a sequência *Salesman*, a projeção do ponto no eixo das abcissas equivale a *taxa de bits* no valor de 768 kbps.

É interessante ao reparar que os valores são de certa forma semelhantes, sendo possível dizer que existe um desempenho maior para *taxas de bits* maiores que 750 kbps. Deve-se lembrar, no entanto, na grande diferença registrada para a complexidade computacional, na qual a *codec* do padrão H.265/HEVC demonstrou-se muito mais custosa computacionalmente, obtendo valores para o tempo de execução de, no mínimo, 3 vezes maiores que o a *codec* do padrão H.264/AVC.

Sendo assim, é perceptível que ambos os padrões apresentam vantagens e desvantagens particulares, que devem ser levadas em conta ao implementar um serviço de transmissão/reprodução de vídeo digital. Visto que os serviços de vídeo são heterogêneos em relação aos seus requisitos, cabe um estudo de caso para avaliar em termos práticos a efetividade de cada um dos padrões.

## 6.6 DISCUSSÃO DA EFETIVIDADE DAS MEDIÇÕES EFETUADAS E MÉTRICAS

Em um ambiente heterogêneo de comunicação, diferentes técnicas de compressão e diferentes *taxas de bits* podem ser utilizados simultaneamente, e, da mesma forma, diferentes tipos de erros pode acontecer de acordo com as diferentes demandas por Qualidade de Serviço (QoS). Devido a sua simplicidade, o Erro Quadrático Médio (MSE) e a Razão Sinal-Ruído de Pico (PSNR) são métricas de qualidade amplamente utilizadas. No entanto, geralmente, estas não são capazes de fornecer uma medida objetiva sobre qualidade correspondente à qualidade percebida por um observador humano para uma ampla gama de parâmetros de codificação e transmissão (VRANJEŠ, RIMAC-DRLJE, ŽAGAR, 2007, tradução nossa).

Já a maior parte de estudos analíticos da complexidade de codificadores de vídeo, prioriza em testar cada parâmetro independentemente, comparando o desempenho de uma configuração padrão contra a mesma configuração, tendo uma ferramenta ativada de cada vez. No entanto, algoritmos atuais de codificação de vídeo são caracterizados por altos níveis de interdependência entre ferramentas de codificação, o que significa que qualquer ganho adicional obtido ativando uma opção de codificação específica, pode ser dependente da condição de ativado ou desativado de outras ferramentas de codificação (CORREA et al, 2012, tradução nossa).

## 7 CONCLUSÃO

A linha de pesquisa de codificação de vídeo em Computação gráfica, é um campo em constante aperfeiçoamento e expansão. Cada vez mais se faz necessário o uso de tecnologias sofisticadas e eficientes no tratamento de vídeos digitais. A padronização H.265/HEVC surge como uma solução para um futuro próximo, no qual teremos a transmissão e reprodução de vídeos digitais com resoluções Ultra HD.

Por meio da pesquisa conduzida neste trabalho foi possível avaliar a efetividade dos padrões analisados e comprovar a superioridade em desempenho do padrão H.265/HEVC em relação ao padrão H.264/AVC, com ênfase para taxa de bits maiores que 750kbps. No entanto, constatou-se também, que a complexidade computacional demonstrou-se, no mínimo, na ordem de 3 vezes maior.

Ficou evidenciado que o padrão H.265/HEVC tem maior ganho para sequências com intensidade de movimentos, tal como na sequência *Foreman*, a qual possui movimentos de câmera, o que amplia a dificuldade de estimação e compensação de movimentos no processo de codificação. No entanto, atualmente, o H.264/AVC é uma solução viável para comunicações de vídeo que vão desde videoconferências à difusão de sinal de TV de alta definição em larga escala.

Para trabalhos futuros, sugere-se o emprego de sequências de vídeo de maiores resoluções nos testes, incluindo as Ultra resoluções (4k e 8k). Sugere-se, também, um estudo mais aprofundado da engenharia por trás do projeto do padrão H.265/HEVC, e, o uso de técnicas e ferramentas de processamento paralelo para tirar proveito das capacidades de sua arquitetura.

## REFERÊNCIAS

- BINDAL, Shivam; KHANNA, Udit; SHARMA, Manoj. Trends in Video Compression Technologies and Detailed Performance Comparison of H.264/MPEG-AVC and H.265/MPEG-HEVC. In: **International Journal of Engineering Research & Technology (IJERT)**, New Delhi, India, v.3, n.12, p.748-754, dec. 2014. Disponível em: <<http://www.ijert.org/download/12063/trends-in-video-compression-technologies-and-detailed-performance-comparison-of-h264mpeg-avc-and-h265mpeg-hevc>> Acesso em: 30 mar. 2015.
- CORREA, Guilherme; ASSUNCAO, Pedro; AGOSTINI, Luciano; CRUZ, Luis A. da Silva. Performance and Computational Complexity Assessment of High Efficiency Video Encoders. In: **Circuits and Systems for Video Technology, IEEE Transactions on**, v.22, n.12, p. 1899-1909, out. 2012. Disponível em: <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6327641>> Acesso em: 14 out. 2015.
- DIAS, Thiago Vaz de A. C. ; ALCAIM, Abraham. Codificação de vídeo em H.264. In: XIV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA PUC-RIO, 2006, PUC-Rio. **Anais...** Local da publicação: PUC-rio, 2006. Disponível em: <[http://www.puc-rio.br/pibic/relatorio\\_resumo2006/relatorio/CTC/Cetuc/Thiago%20Vaz%20de%20A.%20C.%20Dias.pdf](http://www.puc-rio.br/pibic/relatorio_resumo2006/relatorio/CTC/Cetuc/Thiago%20Vaz%20de%20A.%20C.%20Dias.pdf)> Acesso em: 17 mar. 2015.
- GROIS, Dan; MARPE, Detlev; MULAYOFF, Amit; ITZHAKY, Benaya; HADAR, Ofer. Performance Comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders. In: PICTURE CODING SYMPOSIUM (PCS), 30., 2013, California. **Anais...** San Jose: IEEE, 2013, p. 4. Disponível em: <[http://iphome.hhi.de/marpe/download/Performance\\_HEVC\\_VP9\\_x264\\_PCS\\_2013\\_preprint.pdf](http://iphome.hhi.de/marpe/download/Performance_HEVC_VP9_x264_PCS_2013_preprint.pdf)> Acesso em: 31 out. 2015.
- GROIS, Dan; MARPE, Detlev; NGUYEN, Tung; HADAR, Ofer. Comparative assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders for low-delay video applications. In: **Applications of Digital Image Processing XXXVII**, San Diego – CA, USA, v. 9217, set. 2014. Disponível em: <<http://dx.doi.org/10.1117/12.2073323>> 05 nov. 2015.

**H.264.2: Reference software for ITU-T H.264 advanced video coding.** Disponível em: <<http://www.itu.int/rec/T-REC-H.264.2>> Acesso em: 06 dez. 2015.

**H.265.2: Reference software for ITU-T H.265 high efficiency video coding.** Disponível em: <<http://www.itu.int/rec/T-REC-H.265.2>> Acesso em: 06 dez. 2015.

HUANGYUAN, Qingxiong; SONG, Li; LUO, Zhengyi; WANG, Xiangwen; ZHAO, Yanan. Performance Evaluation of H.265/MPEG-HEVC Encoders for 4K Video Sequences. In: 2014 ANNUAL SUMMIT AND CONFERENCE (APSIPA) - Shanghai, China. **Anais...** ASIA-PACIFIC SIGNAL AND INFORMATION PROCESSING ASSOCIATION, dez. 2014. p. 7. Disponível em: <[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=7041782&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7041782&tag=1)> Acesso em: 05 nov. 2015.

International Telecommunication Union. **Recommendation ITU-T H.264.** feb. 2014. Suíça: Telecommunication Standardization Sector of ITU, 2014.

International Telecommunication Union. **Recommendation ITU-T H.265.** out. 2014. Suíça: Telecommunication Standardization Sector of ITU, 2014.

KALPANA, Seshadrinathan; SOUNDARARAJAN, Rajiv; BOVIK, Alan C.; CORMACK, Lawrence K. Study of Subjective and Objective Quality Assessment of Video. In: **IEEE Transactions on Image Processing**, New Jersey, USA, v.19, n.6, p.1427-1441, jun. 2009. Disponível em: <[http://live.ece.utexas.edu/publications/2010/ks\\_tip\\_june10.pdf](http://live.ece.utexas.edu/publications/2010/ks_tip_june10.pdf)> Acesso em: 01 nov. 2015.

KOUMARAS, Harilaos; KOURTIS, Michail-Alexandros; MANTZOURATOS, Spyros; MARTAKOS, Drakoulis. Quantitative Performance Evaluation Of the Emerging HEVC/H.265 Video Codec. In: QOEMCS 2012 WORKSHOP, EURO ITV 2012, 06 July 2012, Berlin – Alemanha. **Anais...** QoEMCS 2012 workshop, 2012. p. 4. Disponível em: <<http://aias.iit.demokritos.gr/~koumaras/EuroITV%20KOUMARAS-%20Camera%20ready.pdf>> Acesso em: 17 mar. 2015.

LAZZARIN, Luiz Felipe. **Distribuição de vídeo escalável baseada na recomendação H.264/SVC em um serviço de WEBTV**. 2014. 91f. TCC (Bacharel em Ciência da Computação) – UNESC, Criciúma – SC, Brasil.

MEDEIROS, Tiago Pizzetti. **Segmentação Quadtree para Seleção de blocos na codificação de vídeo H.264/AVC**. 2011. 105 f. TCC (Bacharel em Ciência da Computação) – UNESC, Criciúma – SC, Brasil.

MARPE, Detlev; WIEGAND, Thomas; SULLIVAN, Gary J.. **The H.264/MPEG4 Advanced Video Coding Standard and its Applications**. Communications Magazine IEEE, v. 4, n. 8, p. 134-143, ago. 2006. Disponível em: <<http://iphome.hhi.de/wiegand/assets/pdfs/h264-AVC-Standard.pdf>> Acesso em: 22 de abr. 2015.

MARPE, Detlev; GORDON, Stephen; WIEGAND, Thomas. H.264/MPEG4-AVC Fidelity Range Extensions: Tools, Profiles, Performance, and Application Areas. In: **Proc. ICIP**, Genova, Italy, Sept. 2005. Disponível em: <[http://iphome.hhi.de/marpe/download/ICIP05\\_H264\\_Fidelity.pdf](http://iphome.hhi.de/marpe/download/ICIP05_H264_Fidelity.pdf)> Acesso em: 05 nov. 2015.

MEDEIROS, Tiago Pizzetti. **Segmentação Quadtree para Seleção de blocos na codificação de vídeo H.264/AVC**. 2011. 105 f. TCC (Bacharel em Ciência da Computação) – UNESC, Criciúma – SC, Brasil.

MILOVANOVIC, Dragorad; BOJKOVIC, Zoran. MPEG Video deployment in interactive multimedia systems: HEVC vs. AVC codec performance study. In: WSEAS TRANSACTIONS ON SIGNAL PROCESSING, 2013, University of Belgrade. **Anais...** Belgrado, Sérvia: University of Belgrade, 2013. p. 167-176. Disponível em: <<http://www.wseas.org/multimedia/journals/signal/2013/085702-235.pdf>> Acesso em: 17 de mar. 2015.

OHM, Jens-Rainer; SULLIVAN, Gary J.; SCHWARZ, Heiko; TAN, Thiew Keng; WIEGAND, Thomas. Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC). In: **IEEE Transactions on Circuits and Systems for Video Technology**, v. 22, n. 12, p. 1669-1684, dec. 2013. Disponível em: <

Coding\_WS12/selected-readings/2012\_12\_IEEE-HEVC-Performance.pdf> Acesso em: 05 nov. 2015.

PINSON, Margaret H.; WOLF, Stephen. A New Standardized Method for Objectively Measuring Video Quality. In: **IEEE Transactions on Broadcasting**, Boulder - CO, USA, nov. 2004. Disponível em:

<[http://atc.umh.es/gatcom/bin/oqam/Referencias/Pinson\\_Wolf\\_2004.pdf](http://atc.umh.es/gatcom/bin/oqam/Referencias/Pinson_Wolf_2004.pdf)> Acesso em: 05 nov. 2015.

SEELING, Patrick; REISSLEIN, Martin. Video Traffic Characteristics of Modern Encoding Standards: H.264/AVC with SVC and MVC Extensions and H.265/HEVC. **The Scientific World Journal**, 2014. vol. 2014, p. 17. Disponível em:

<<http://downloads.hindawi.com/journals/tswj/2014/189481.pdf>> Acesso em: 17 de mar. 2015.

SCHWARZ, Heiko; MARPE, Detlev; WIEGAND, Thomas. **Overview of the Scalable Video Coding Extension of the H.264/AVC Standard**. IEEE Transactions on Circuits and Systems for Video Technology, v. 17, n. 9, p.1103-1120, set. 2007.

Disponível em:

<[http://ip.hhi.de/imagecom\\_G1/assets/pdfs/Overview\\_SVC\\_IEEE07.pdf](http://ip.hhi.de/imagecom_G1/assets/pdfs/Overview_SVC_IEEE07.pdf)> Acesso em: 06 mai. 2015.

SILVA, André Marcelo Coelho da. **Um Estudo Sobre o Padrão H.264/AVC de Compressão de Vídeo**. 2007. 44 f. Monografia (Especialização) - Curso de Pós-graduação em Informática, Universidade Católica de Pelotas, Pelotas. Disponível em: <<http://ppginf.ucpel.tche.br/TI-arquivos/2007/PPGINF-UCPel-TI-2007-2-02.pdf>> Acesso em: 31 out. 2015.

SILVA, Isis Vanielle Gomes da. **Análise comparativa entre os codificadores de vídeo H.264/AVC e H.265/HEVC**. 2014. 91 f. TCC (Bacharel em Engenharia de Telecomunicações) – Universidade de Pernambuco, Escola Politécnica de Pernambuco, Pernambuco, Brasil.

SOUZA, Marlon Henrique de. **Avaliação de desempenho da escalabilidade espacial na recomendação H.264/SVC para codificação de vídeo**. 2011. 72 f. TCC (Bacharel em Ciência da Computação) – UNESC, Criciúma – SC, Brasil.

SULLIVAN, Gary J.; OHM, Jens-Rainer; HAN, Woo-Jin; WIEGAND, Thomas. Overview of the High Efficiency Video Coding (HEVC) Standard. **IEEE Transactions on circuits and systems for video technology**, v.22, n.12, p.1649 – 1668, set./dez. 2012. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6316136>> Acesso em: 30 de mar. 2015.

SULLIVAN, Gary J.; WIEGAND, Thomas. Video Compression—From Concepts to the H.264/AVC Standard. In: PROCEEDINGS OF THE IEEE, 2005. **Anais... IEEE**, 2005. p. 18-31. Disponível em: <[http://ip.hhi.de/imagecom\\_G1/assets/pdfs/pieee\\_sullivan\\_wiegand\\_2005.pdf](http://ip.hhi.de/imagecom_G1/assets/pdfs/pieee_sullivan_wiegand_2005.pdf)> Acesso em: 17 març. 2015.

SULLIVAN, Gary J.; BOYCE, Jill M.; CHEN, Ying; OHM, Jens-Rainer. SEGALL, C.Andrew; VETRO, Anthony. Standardized Extensions of High Efficiency Video Coding (HEVC). In: IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, dez. 2013. **Anais... IEEE**, 2013. p. 1001-1016. Disponível em: <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6630053>> Acesso em: 05 nov. 2015.

SZE, Vivienne ; BUDAGAVI, Madhukar. A Comparison of CABAC throughput for HEVC/H.265 vs. AVC/H.264. In: **2013 IEEE Workshop on Signal Processing Systems**, 2013. Disponível em: <[http://www.rle.mit.edu/eems/wp-content/uploads/2013/11/sze\\_sips\\_2013.pdf](http://www.rle.mit.edu/eems/wp-content/uploads/2013/11/sze_sips_2013.pdf)> Acesso em: 05 nov. 2015.

**VLC media player**. VideoLan. Disponível em: <<http://www.videolan.org/vlc/>> Acesso em: 09 nov. 2015.

VRANJEŠ, M.; RIMAC-DRLJE, S.; ŽAGAR, D. Objective Video Quality Metrics. In: **ELMAR**, Osijek, Croácia, 2007. Disponível em: <[http://www.researchgate.net/publication/4305738\\_Objective\\_video\\_quality\\_metrics](http://www.researchgate.net/publication/4305738_Objective_video_quality_metrics)> Acesso em: 04 nov. 2015.

**YUV Video Sequences**. Arizona State University. Disponível em: <<http://trace.eas.asu.edu/>> Acesso em: 09 nov. 2015.

**YUV Player Deluxe.** Disponível em: <<http://yuvplayer.com/index.html>> Acesso em:  
09 nov. 2015.

## APÊNDICE A - ARTIGO

# ANÁLISE COMPARATIVA ENTRE OS PADRÕES ITU-T H.265/HEVC E H.264/AVC EM CODIFICAÇÃO DE VÍDEO

Allan Serra Braga Bugyi<sup>1</sup>, Evânio Ramos Nicoleit<sup>2</sup>

<sup>1</sup>Acadêmico do curso de Ciência da Computação – Unidade Acadêmica de Ciências, Engenharias e Tecnologias - Universidade do Extremo Sul Catarinense (UNESC) - Criciúma - SC

<sup>2</sup>Professor do curso de Ciência da Computação - Unidade Acadêmica de Ciências, Engenharias e Tecnologias - Universidade do Extremo Sul Catarinense (UNESC) - Criciúma – SC

allan.sb2@hotmail.com, ern@unesc.net

**Abstract.** *The current paper performs a objective comparative analysis in video coding. The subjects of study are the standards created by International Telecommunication Union – Telecommunication Sector (ITU-T) and ISO/IEC Moving Pictures Experts Group (MPEG), known as H.264, or, Advanced Video Coding (AVC), and, H.265, or, High Efficient Video Coding (HEVC). The development of this video coding standard focus to attend future needs placed on the transmission ways and diverse video-reproducer devices, based on a demand ever increasing for quality and diminished bitrate. Such standardization was designed in a way to be prepared to handle Ultra HD resolutions, known as 4k and 8k. The results show the efficiency of H.265/HEVC standard, with emphasis for greater bitrates than 750 kbps. However, the computational cost is still high, with 3 times higher, at least, relative to H.264/AVC.*

**Resumo.** *O presente trabalho realiza análise comparativa do tipo objetiva em codificação de vídeo. São objetos de estudo os padrões criados pelas organizações, International Telecommunication Union – Telecommunication Sector (ITU-T) e ISO/IEC Moving Pictures Experts Group (MPEG), conhecidos como H.264, ou, Codificação de Vídeo Avançada (AVC), e, H.265, ou, Codificação de Vídeo de Alta Eficiência (HEVC). O desenvolvimento deste recente padrão em codificação de vídeo objetiva atender futuros desafios impostos sobre as redes de comunicação e dispositivos diversos reprodutores de vídeo, baseado numa demanda sempre crescente por qualidade e de reduzida taxa de bits. Tal padronização foi projetada de forma a estar preparada para lidar com resoluções Ultra HD, resoluções conhecidas como 4k e 8k. Os resultados demonstram a eficiência do padrão H.265/HEVC, principalmente para taxa de bits maiores que 750 kbps. Contudo, o custo computacional ainda é elevado, em pelo menos 3 vezes mais em relação ao padrão H.264/AVC.*

## 1. Introdução

Na última década e meia, tecnologias de compressão de vídeo digital se tornaram uma parte fundamental do modo como se cria, comunica e se consome informação visual. Os avanços no desempenho das codificações de vídeo devem-se a implementação de diferenciadas abordagens e estratégias que buscam a otimização computacional para vídeos digitais (SULLIVAN; WIEGAND, 2005, tradução nossa).

Atualmente, é comum o uso do termo 'alta definição' (HD), referindo-se a serviços e dispositivos que suportam resoluções iguais ou superiores a 720 pixels. Um crescente número de serviços de streaming e de vídeo sob demanda (VoD) tem disponibilizado amplamente conteúdo de alta definição. Não obstante, o mercado de dispositivos móveis tem se difundido globalmente a passos largos, oferecendo dispositivos para a população em geral, capazes de reproduzir e criar conteúdo visual em qualidade equivalente.

Além disso, a busca por qualidades superiores ao reproduzir vídeos digitais, é uma tendência que futuramente criará a necessidade de suplantando as resoluções atuais por outras ainda maiores, como as resoluções 4k e 8k. Como efeito decorrente desta demanda, impõe-se um custo computacional maior para sistemas de compressão e reprodução de vídeos digitais, assim como desafios para os meios de transmissão (SULLIVAN et al, 2012, tradução nossa).

Enquadra-se dentro do escopo deste trabalho, uma análise comparativa do tipo objetiva entre os padrões de codificação da ITU-T e ISO-IEC H.264/AVC e H.265/HEVC. Tal empreitada se deu por meio de testes de codificação realizados com sequências de vídeo em formato planar. Durante os testes foram armazenados as informações resultantes do processo de codificação, de forma que após uma cuidadosa análise de tais tomou-se lugar uma comparação baseada em métricas quantitativas amplamente empregadas em pesquisas em codificação de vídeo – Sinal-Ruído de Pico (PSNR), taxa de bits e tempo de execução. A partir destas, foi possível avaliar o desempenho e a complexidade computacional das sequências de vídeo codificadas.

## 2. Codificação de vídeo e padrões ITU-T/ISO-IEC

Para a codificação de vídeo, define-se os seguintes objetivos: transmitir vídeos usando a menor taxa de bits possível, mantendo-se a fidelidade de reprodução em comparativo ao vídeo original. Um sistema de codificação que cumpre com tais expectativas, possui: eficiência de codificação e taxas de distorção aceitáveis em relação ao vídeo original. Tal sistema pode ser chamado de *codec*, ou seja, um sistema que engloba um codificador e um decodificador (SULLIVAN; WIEGAND, 2005, tradução nossa).

O padrão de Codificação de Vídeo Avançada (AVC), ou simplesmente H.264 – publicado no ano de 2003 -, está atualmente em vigor sendo utilizado em diversos tipos de serviços de vídeo, como televisão digital, cinema digital, vídeo teleconferência, vídeos da Internet, serviços de compartilhamento de vídeo – como Youtube, por exemplo -, Vídeo sob Demanda (*Video on Demand* - VoD), vídeos caseiros/amaradores e discos Blu-ray.

O processo de codificação descrito por este padrão faz uso de modelos matemáticos para predição e otimização da informação visual presente num vídeo digital. A sua arquitetura está dividida em duas camadas, uma de rede – Camada de

Abstração de rede, responsável pela formatação e envio dos dados pelos meios de transmissão - e uma camada responsável por lidar com a codificação de vídeo propriamente dita – Camada de Codificação de Vídeo. Ambas possuem vários níveis de abstração e interdependência de dados que visam robustez e a resiliência a erros.

Sendo previsível que o padrão H.264/AVC também terá um prazo de efetividade, a ITU-T Video Coding Experts Group (VCEG) lançou-se no desenvolvimento de um novo padrão, junto ao grupo ISO/IEC Moving Pictures Experts Group (MPEG), criando o Joint Collaborative Team (JCT) (SULLIVAN; WIEGAND, 2005, tradução nossa).

O padrão de Codificação de Vídeo de Alta Eficiência (HEVC), ou simplesmente H.265 – publicado no ano de 2013 -, toma como base o seu antecessor, mantendo diversos elementos estruturais e implementando novas abordagens para o processo de codificação, como estruturas de dados em forma de árvores quaternárias, unidades de codificação, blocos de codificação, unidades transformadas, blocos transformados, unidades de predição, blocos de predição, entre outras, além de maiores opções para a segmentação dos quadros em blocos e/ou fatias, mais direções para a predição *intra* e arquiteturas de processamento paralelo (SULLIVAN et al, 2012, tradução nossa).

### **3. Análise comparativa entre os padrões ITU-T H.265/HEVC e H.264/AVC em codificação de vídeo**

O presente trabalho propôs-se a efetuar uma análise comparativa de desempenho e complexidade computacional do tipo objetiva-quantitativa entre os dois padrões mais recentes em codificação de vídeo desenvolvidos pelo setor de padronização da *International Telecommunication Union* (ITU-T), e pelo *ISO/IEC Moving Pictures Experts Group* (MPEG).

Foi feito uso de métricas quantitativas, por meio das quais foram efetuadas medições durante e após o processo de codificação. Estas são: PSNR-Y (dB), taxa de bits (kbps) e tempo de execução (s).

#### **3.2 Sequências de vídeo utilizadas**

As sequências utilizadas nos testes caracterizam-se por estarem em formato original, ou seja, um formato planar sem compressão, no caso, YUV.

Foram utilizadas sequências com resoluções de 176 x 144 *pixels*, também conhecido como formato “qcif”. A tabela 1, lista os arquivos utilizados, sua *resolução*, *sub-amostragem de cores* e *extensão*.

**Tabela 1 – Lista de seqüências utilizadas nos testes e suas características.**

<b>Nome da seqüência</b>	<b>Resolução</b>	<b>Sub-amostragem</b>	<b>Classificação</b>	<b>Formato</b>
foreman	176x144	4:2:0	qcif	YUV
miss-america	176x144	4:2:0	qcif	YUV
salesman	176x144	4:2:0	qcif	YUV

A seqüência Foreman caracteriza-se por ter um alto grau de movimentação tanto de fundo como de frente ao longo de seus quadros. Esta é uma seqüência que exige do codificador considerável quantidade de processamento destinado a predição Inter.

Já a seqüência Miss-américa possui um fundo preto e fixo, tendo à frente pequenas mudanças de movimentação referentes à pessoa que está falando. Fazendo-se uso desta seqüência, é possível testar nos codificadores a capacidade de tirar vantagem de uma área relativamente grande, a qual é estática e de pouca variação cromática.

Finalmente, a seqüência *Salesman*, possui um fundo fixo com movimentação de frente e sombra, além de considerável quantidade de detalhes na cena. Tal seqüência permite testar o nível de granularidade resultante do processo de codificação codificadores em relação as diferentes cores e texturas presentes.

### **3.3 Implementações utilizadas**

Buscou-se na Internet implementações dos padrões em análise e foram encontradas implementações da própria ITU-T, as quais foram desenvolvidas como modelos de referência – *JM Reference Model* (H.264/AVC) e *HM Reference Model* (H.265/HEVC). No entanto, estas implementações procuram empregar todas as técnicas de codificação de vídeo elucidadas nos capítulos anteriores, de acordo com a norma de cada padrão.

Sendo estas implementações desenvolvidas pelo próprio órgão padronizador e livres de interesse comercial ou, de customização de terceiros, acreditou-se serem as melhores opções de softwares existentes para se efetuar uma análise imparcial dos padrões.

### **3.4 Parâmetros de execução e arquivos de configuração**

A sintaxe dos comandos e os parâmetros de execução são detalhados na documentação que vem juntamente com os códigos-fonte.

Arquivos de configuração padrão – tanto para o codificador como para o decodificador - também são fornecidos para cada software. Dentro destes arquivos estão especificados todos os parâmetros de execução em forma de texto. Foram mantidos a maioria dos valores já existentes, fazendo-se as alterações necessárias para cada arquivo de entrada e *taxa de bits alvo*.

Foi definido para todos os testes:

- a) *profundidade em bits* para cada pixel de cor no valor de 8 bits;
- b) *taxa de quadros* no valor de 30 quadros por segundo;
- c) *período de codificação Intra* a cada 32 quadros.

Já para os perfis e níveis de cada um dos softwares, foi utilizada a seguinte configuração:

a) *JM Reference Model* – Perfil: *Main* ; nível: 4.0;

b) *HM Reference Model* – Perfil: *Random Access Main* ; nível: 3.1.

Os parâmetros de execução apresentados anteriormente (tabelas 2 e 3) referem-se ao teste efetuado com a *taxa de bits alvo* no valor de 32 kbps, o equivalente a 32.000 bps, como especificado. Abaixo, na tabela 2 lista-se todos os testes realizados.

**Tabela 2 – Testes e taxas de bits alvo utilizadas**

Testes	Taxa de bits alvo					
	1	2	3	4	5	6
<b>JM Reference Model</b>	32 kbps	64 kbps	128 kbps	256 kbps	512 kbps	1024 kbps
<b>HM Reference Model</b>	32 kbps	64 kbps	128 kbps	256 kbps	512 kbps	1024 kbps

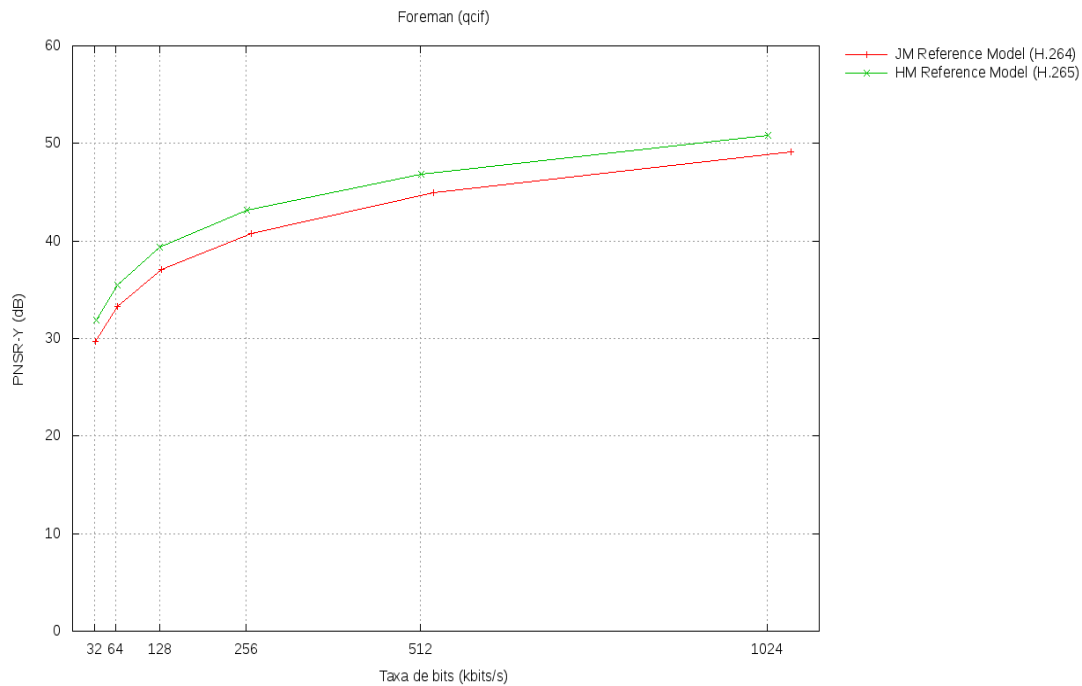
#### 4. Resultados

Os resultados obtidos do processo da codificação das sequências anteriormente mencionadas foram representados em dois tipos de gráficos, de desempenho e complexidade computacional.

O gráfico de *PSNR-Y* em função da *taxa de bits*, relaciona o aumento da qualidade de um vídeo baseado na largura do canal de dados, o qual proporciona melhores resultados quanto mais elevado for. Portanto, a *taxa de bits* influencia diretamente na *taxa de compressão* do arquivo de vídeo original. Taxas de bits menores forçam as *codecs* a adaptarem o fluxo de dados de uma sequência de vídeo, descartando mais informações relativamente a taxas-alvo de bits maiores.

O gráfico de *complexidade computacional* em função da *taxa de bits*, relaciona o custo computacional durante o processo de codificação em termos de unidades de tempo – no caso, segundos (s) – baseado, também, na largura do canal de dados. A *complexidade computacional* tende a se elevar proporcionalmente ao aumento das taxas-alvo de bits, devido ao aumento de dados disponíveis a serem manipulados – aqui refere-se a aplicação de técnicas de compressão de sinal digital: quantização, entropia e transformadas - e preditos.

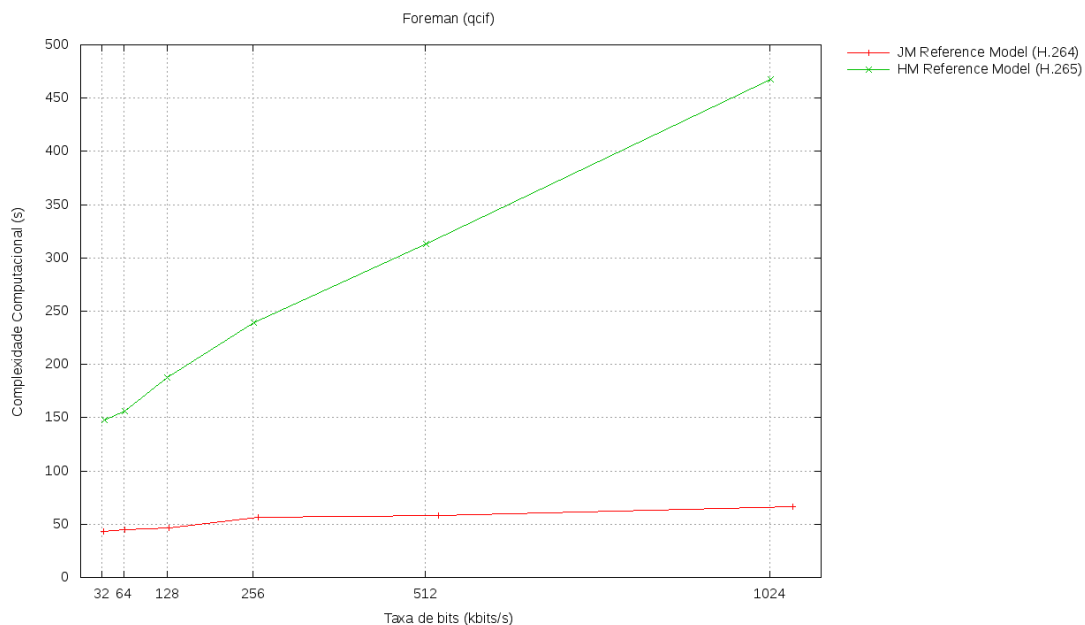
Na figura 4, é possível observar o desempenho de ambos os modelos de referência para a sequência *Foreman*.



**Figura 4. Gráfico PSNR-Y (db) x Taxa de Bits (kbps) para a sequência Foreman**

Ambos os modelos de referência são representados no gráfico em escala logarítmica. É possível ver que o modelo de referência do padrão H.265/HEVC é superior em desempenho e qualidade.

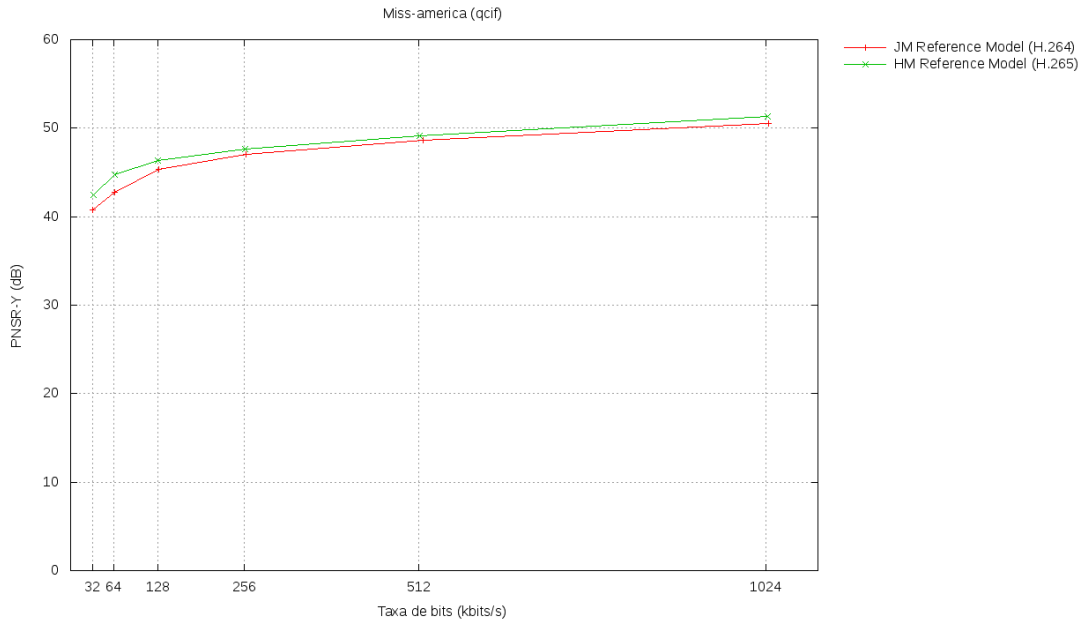
Já na figura 5, é possível observar o custo computacional de ambos os modelos de referência para a sequência Foreman.



**Figura 5. Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Foreman**

Percebe-se que a complexidade computacional do modelo de referência do padrão H.265/HEVC foi consideravelmente maior, apresentando-se com um crescimento constante retilíneo. Por outro lado, o modelo de referência do padrão H.264/AVC, apresentou, relativamente, valores e crescimento consideravelmente menores.

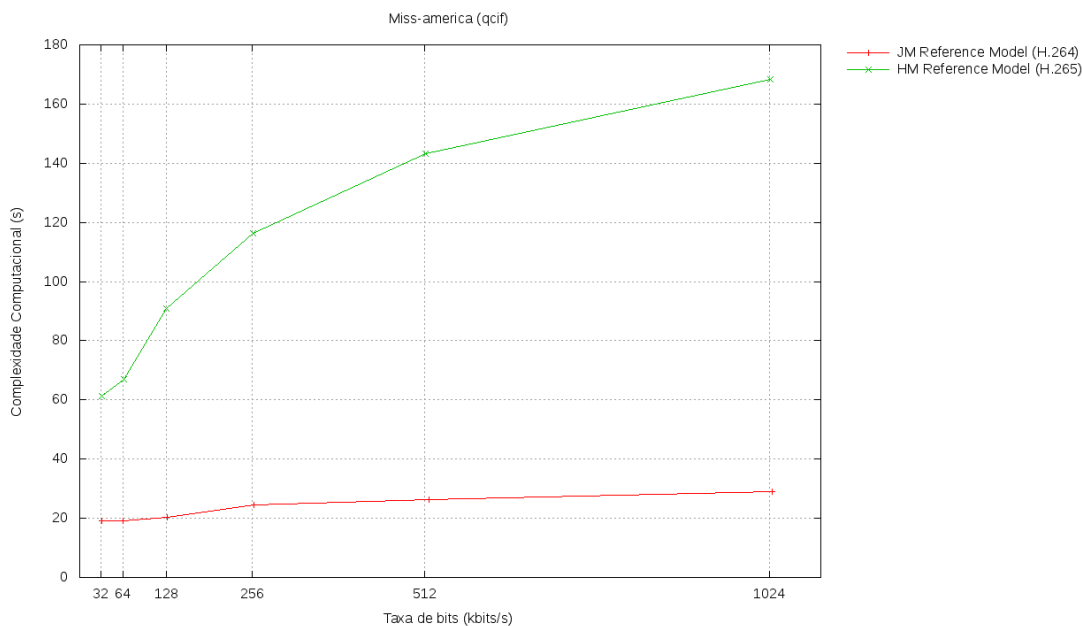
Na figura 6, é possível observar o desempenho de ambos os modelos de referência para a sequência *Miss-américa*.



**Figura 6. Gráfico PSNR-Y x Taxa de Bits para a sequência Miss-américa**

Ambos os modelos de referência são representados no gráfico em escala logarítmica. É possível ver que o modelo de referência do padrão H.265/HEVC é, novamente, superior em desempenho e qualidade, embora que, para esta sequência em específico, os modelos alcançam valores bem próximos. Acredita-se que tal resultado deve-se a grandes porções estáticas e de mesma coloração nos quadros ao longo sequência.

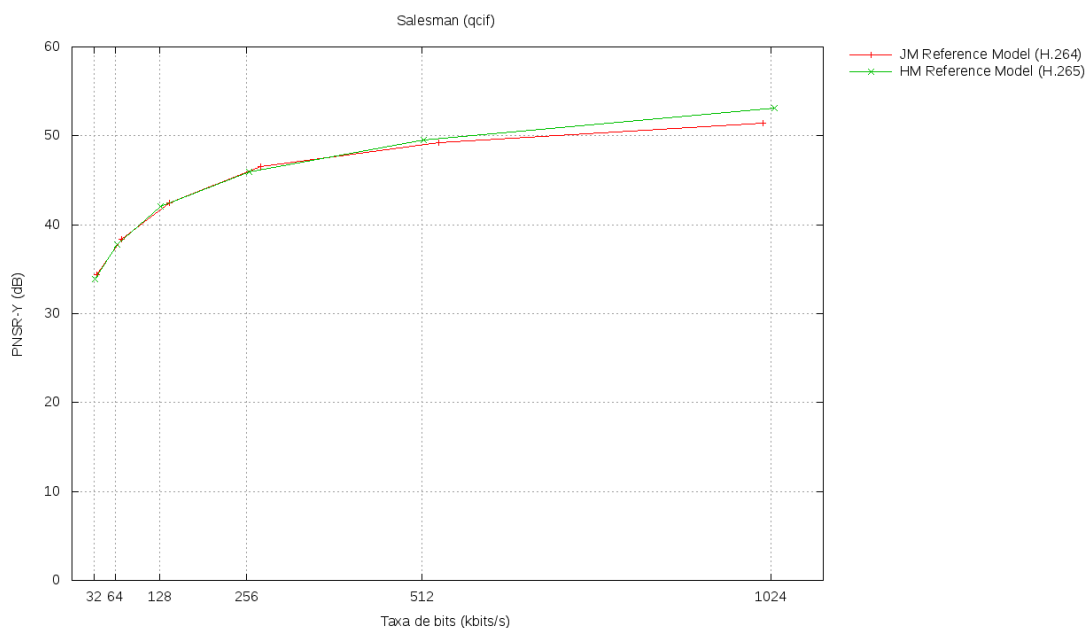
Já na figura 7, é possível observar a complexidade computacional de ambos os modelos de referência para a sequência *Miss-américa*.



**Figura 7. Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Miss-américa**

É possível ver que a complexidade computacional do modelo de referência do padrão H.265/HEVC foi consideravelmente maior, apresentando-se com um crescimento variante e decrescente. Por outro lado, o modelo de referência do padrão H.264/AVC, novamente apresentou, relativamente, valores e crescimento consideravelmente menores.

Na figura 8, é possível observar o desempenho de ambos os modelos de referência para a sequência Salesman.

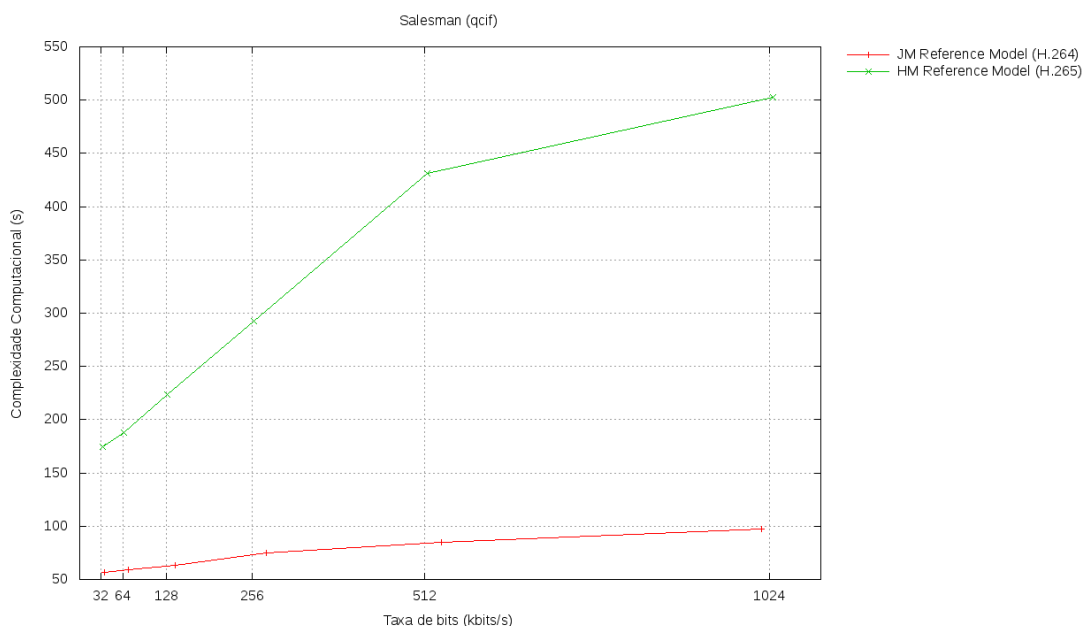


**Figura 8. Gráfico PSNR-Y x Taxa de Bits para a sequência Salesman**

Ambos os modelos de referência são representados no gráfico em escala logarítmica. Desta vez, é possível ver que o modelo de referência do padrão

H.265/HEVC é superior em desempenho e qualidade, apenas após a taxa de bits alcançar um valor mediano entre 256 kbps e 512 kbps. O modelo de referência do padrão H.264/AVC, praticamente equipara-se até este ponto em desempenho e qualidade, tendo uma leve desvantagem próximo a taxa de 256 kbps. No entanto, mais uma vez, o padrão H.265/HEVC se mostra superior para taxas de bits maiores.

Já na figura 9, é possível observar a complexidade computacional de ambos os modelos de referência para a sequência Salesman.



**Figura 9. Gráfico Complexidade Computacional (s) x Taxa de bits (kbps) para a sequência Salesman**

É visível que a complexidade computacional do modelo de referência do padrão H.265/HEVC foi, uma vez mais, consideravelmente maior, apresentando-se com um crescimento variante e decrescente. Por outro lado, o modelo de referência do padrão H.264/AVC, novamente apresentou, relativamente, valores e crescimento consideravelmente menores.

## 5. Conclusão

A linha de pesquisa de codificação de vídeo em Computação gráfica, é um campo em constante aperfeiçoamento e expansão. Cada vez mais se faz necessário o uso de tecnologias sofisticadas e eficientes no tratamento de vídeos digitais. A padronização H.265/HEVC surge como uma solução para um futuro próximo, no qual teremos a transmissão e reprodução de vídeos digitais com resoluções Ultra HD.

Por meio da pesquisa conduzida neste trabalho foi possível avaliar a efetividade dos padrões analisados e comprovar a superioridade em desempenho do padrão H.265/HEVC em relação ao padrão H.264/AVC, com ênfase para taxa de bits maiores que 750kbps. No entanto, constatou-se também, que a complexidade computacional demonstrou-se, no mínimo, na ordem de 3 vezes maior.

Ficou evidenciado que o padrão H.265/HEVC tem maior ganho para sequências com intensidade de movimentos, tal como na sequência Foreman, a qual possui movimentos de câmera, o que amplia a dificuldade de estimação e compensação de

movimentos no processo de codificação. No entanto, atualmente, o H.264/AVC é uma solução viável para comunicações de vídeo que vão desde video-conferências à difusão de sinal de TV de alta definição em larga escala.

Para trabalhos futuros, sugere-se o emprego de sequências de vídeo de maiores resoluções nos testes, incluindo as Ultra resoluções (4k e 8k). Sugere-se, também, um estudo mais aprofundado da engenharia por trás do projeto do padrão H.265/HEVC, e, o uso de técnicas e ferramentas de processamento paralelo para tirar proveito das capacidades de sua arquitetura.

## 6. Referências

DIAS, Thiago Vaz de A. C. ; ALCAIM, Abraham. Codificação de vídeo em H.264. In: XIV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA PUC-RIO, 2006, PUC-Rio. **Anais...** Local da publicação: PUC-rio, 2006. Disponível em: <[http://www.puc-rio.br/pibic/relatorio\\_resumo2006/relatorio/CTC/Cetuc/Thiago%20Vaz%20de%20A.%20C.%20Dias.pdf](http://www.puc-rio.br/pibic/relatorio_resumo2006/relatorio/CTC/Cetuc/Thiago%20Vaz%20de%20A.%20C.%20Dias.pdf)> Acesso em: 17 mar. 2015.

GROIS, Dan; MARPE, Detlev; NGUYEN, Tung; HADAR, Ofer. Comparative assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders for low-delay video applications. In: **Applications of Digital Image Processing XXXVII**, San Diego – CA, USA, v. 9217, set. 2014. Disponível em: <<http://dx.doi.org/10.1117/12.2073323>> 05 nov. 2015.

**H.264.2: Reference software for ITU-T H.264 advanced video coding.** Disponível em: <<http://www.itu.int/rec/T-REC-H.264.2>> Acesso em: 06 dez. 2015.

**H.265.2: Reference software for ITU-T H.265 high efficiency video coding.** Disponível em: <<http://www.itu.int/rec/T-REC-H.265.2>> Acesso em: 06 dez. 2015.

KOUMARAS, Harilaos; KOURTIS, Michail-Alexandros; MANTZOURATOS, Spyros; MARTAKOS, Drakoulis. Quantitative Performance Evaluation Of the Emerging HEVC/H.265 Video Codec. In: QOEMCS 2012 WORKSHOP, EURO ITV 2012, 06 July 2012, Berlin – Alemanha. **Anais...** QoEMCS 2012 workshop, 2012. p. 4. Disponível em: <<http://aias.iit.demokritos.gr/~koumaras/EuroITV%20KOUMARAS-%20Camera%20ready.pdf>> Acesso em: 17 mar. 2015.

SILVA, Isis Vanielle Gomes da. **Análise comparativa entre os codificadores de vídeo H.264/AVC e H.265/HEVC.** 2014. 91 f. TCC (Bacharel em Engenharia de Telecomunicações) – Universidade de Pernambuco, Escola Politécnica de Pernambuco, Pernambuco, Brasil.

SULLIVAN, Gary J.; OHM, Jens-Rainer; HAN, Woo-Jin; WIEGAND, Thomas. Overview of the High Efficiency Video Coding (HEVC) Standard. **IEEE Transactions on circuits and systems for video technology**, v.22, n.12, p.1649 – 1668, set./dez. 2012. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6316136>> Acesso em: 30 de mar. 2015.

SULLIVAN, Gary J.; WIEGAND, Thomas. Video Compression—From Concepts to the H.264/AVC Standard. In: PROCEEDINGS OF THE IEEE, 2005. **Anais...** IEEE, 2005. p. 18-31. Disponível em: <[http://ip.hhi.de/imagecom\\_G1/assets/pdfs/pieee\\_sullivan\\_wiegand\\_2005.pdf](http://ip.hhi.de/imagecom_G1/assets/pdfs/pieee_sullivan_wiegand_2005.pdf)> Acesso em: 17 març. 2015.