

**JOE LENNON FERREIRA SOARES**

**UM ESTUDO COMPARATIVO DAS ESTRUTURAS DE CODIFICAÇÃO DE  
VIDEOS DOS PADRÕES H.264/AVC E H.266/VVC**

**ALEGRETE**

**2021**



**JOE LENNON FERREIRA SOARES**

**Um Estudo Comparativo Das Estruturas De Codificação De Vídeos Dos Padrões  
H.264/AVC E H.266/VVC**

Trabalho de Conclusão de Curso  
apresentado ao Curso de Engenharia de  
Telecomunicações da Universidade  
Federal do Pampa Campus Alegrete,  
como requisito parcial para obtenção do  
grau de **Bacharel em Engenharia de  
Telecomunicações.**

Orientador: Prof. Bruno Boessio Vizzotto

**ALEGRETE**

**2021**

**JOE LENNON FERREIRA SOARES**

**UM ESTUDO COMPARATIVO DAS ESTRUTURAS DE CODIFICAÇÃO DE VÍDEOS DOS  
PADRÕES H.264/AVC E H.266/VVC**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Telecomunicações da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Telecomunicações.

Trabalho de Conclusão de Curso defendido e aprovado em: 10 de Janeiro de 2022.

Banca examinadora:

---

Prof. Dr. Bruno Boessio Vizzotto

Orientador

UNIPAMPA

---

Prof. Dr. Cristian Muller

UNIPAMPA

---

Prof. Dr. Mauro Fonseca Rodrigues

UNIPAMPA



Assinado eletronicamente por **BRUNO BOESSIO VIZZOTTO, PROFESSOR DO MAGISTERIO SUPERIOR**, em 12/01/2022, às 10:17, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **MAURO FONSECA RODRIGUES, PROFESSOR DO MAGISTERIO SUPERIOR**, em 12/01/2022, às 10:21, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **CRISTIAN MULLER, PROFESSOR DO MAGISTERIO SUPERIOR**, em 12/01/2022, às 12:07, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



A autenticidade deste documento pode ser conferida no site [https://sei.unipampa.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0710208** e o código CRC **16C50E66**.

Universidade Federal do Pampa, Campus Alegrete  
Av. Tiarajú, 810 – Bairro: Ibirapuitã – Alegrete – RS CEP: 97.546-550  
Telefone: (55) 3422-8400



## AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida e pelas oportunidades que Ele me deu durante minha vida, que por diversas vezes eu estivesse abalado, seus planos sempre foram maiores para minha pessoa.

Deixo aqui o meu enorme agradecimento a toda minha família, minhas tias Luciana, Vanuza, Vânia, meus tios José Neto, Willian, Lelton, aos meus avós Luis, Maria de Lourdes e a minha querida e falecida avó Terezinha, que mesmo distante sempre se fizeram presentes, dando-me apoio e incentivo em todas as etapas de minha vida e também da graduação. Destaco também o grande apoio do meu pai José Odilon por sempre se preocupar comigo e me acolher.

Em especial agradeço a minha mãe Lília que não mediu esforços para que esta conquista se concretizasse, sempre me apoiando e incentivando independente das circunstâncias. Com certeza a pessoa mais forte que já conheci e maior inspiração de minhas conquistas.

Sou muito grato aos professores que passaram por minha jornada, por todo o aprendizado e lições passadas durante esta trajetória. Em especial, ao professor Bruno Vizzotto, pela orientação, por não desistir de mim e se fazer presente sempre que solicitado, deixo meus sinceros agradecimentos.

E por fim, agradeço a todos os meus colegas e amigos que fiz durante esta caminhada, sempre lembrarei de vocês e os levarei eternamente. Principalmente ao Gabriel, Bruno, Sinara, Melissa, Magno, Idenilton e Liebert que tive o prazer de compartilhar muitos momentos de alegrias e tristezas, sem vocês eu jamais teria conseguido. Em especial, a minha amada Isadora, pelos momentos compartilhados, palavras de incentivo e carinho que me fizeram perseverar em minha caminhada.

*“Livre para poder sorrir, sim.  
Livre para poder buscar o  
meu lugar ao Sol.”  
(Charlie Brown Jr.)*



## RESUMO

Este trabalho tem como objetivo trazer os conceitos abordados nos padrões de codificação H.264/AVC e H.266/VVC, a fim de se estudar suas estruturas e principais diferenças em seus codificadores. Tratando-se de serviços como *streaming* de vídeo, teleconferência e vídeos com resolução UHD (do inglês, *Ultra High Definition*), os estudos sobre codificadores de vídeo são considerados de vital importância. O padrão H.264/AVC trouxe em seu módulo novas ferramentas na etapa de predição interquadros chamados de estimação e compensação de movimento, os quais foram responsáveis pela maior parte da eficiência de codificação neste padrão. Com os novos vídeos de resolução de 8k-16k, o padrão H.266/VVC obtém uma eficiência estimada em 50% acima de seu antecessor, devido a novas ferramentas implementadas, como estruturas de particionamento flexíveis, novas ferramentas de predição intrablocos e um módulo de filtros *in loop*. Entretanto, o grande avanço das técnicas de codificação inseriu grande complexidade algorítmica sobre as estruturas, resultando em um grande aumento do custo computacional. A partir disto, este estudo evidencia etapas que podem ser exploradas por técnicas que visem reduzir a complexidade computacional destas etapas e assim reduzir o tempo de codificação no novo padrão VVC.

**Palavras-Chave:** Imagens; Vídeo; Compressão de dados; Padrão de Codificação.

## ABSTRACT

This work aims to bring the concepts covered in the encoding standards H.264/AVC and H.266/VVC, in order to study their structures and main differences in their encoders. In terms of services such as video streaming, teleconferencing and videos with UHD resolution (Ultra High Definition), studies on video encoders are considered of vital importance. The H.264/AVC standard brought in its module new tools in the interframe prediction stage called motion estimation and compensation, which were responsible for most of the encoding efficiency in this standard. With the new 8k-16k resolution videos, the H.266/VVC standard achieves an efficiency estimated at 50% above its predecessor, due to new tools implemented, such as flexible partitioning structures, new intra-block prediction tools and a in loop filters module. However, the great advance of coding techniques introduced great algorithmic complexity on the structures, resulting in a great increase in computational cost. From this, this study highlights steps that can be explored by techniques that aim to reduce the computational complexity of these steps and thus reduce the coding time in the new VVC standard.

**Keywords:** Images, Video, Data Compression, Encoding Pattern.



## LISTA DE FIGURAS

Figura 1 - Potenciais melhoras e novos serviços do padrão VVC.....	18
Figura 2 - Sistema de Aquisição de Imagem.....	21
Figura 3 - Níveis Básicos do Processamento de Imagens. ....	23
Figura 4 - a) Imagem Contínua, (b). Uma Linha Escaneada de A para B na Imagem Contínua, (c) Amostragem e Quantização, (d) Linha Digital Escaneada.....	24
Figura 5 - Mesma Imagem com Resoluções Diferentes.....	26
Figura 6 - Matriz Tridimensional $V(m, n, l)$ que Representa um Vídeo Digitalizado. .	27
Figura 7 - Diagrama de Blocos de um Codec Básico.....	28
Figura 8 - Transformada do Cosseno Aplicada a uma Imagem. ....	32
Figura 9 - Modelo de um Codec de Vídeo .....	34
Figura 10 - Linha do Tempo dos Principais Codificadores Desenvolvidos na História. 35	
Figura 11 - Diagrama de blocos do Codec padrão H.264. ....	37
Figura 12 - Processo de estimação de movimento. ....	40
Figura 13 - Comparação entre quadro com e sem Estimação de Movimento.....	41
Figura 14 - Divisão de macroblocos em partições. ....	41
Figura 15 - Divisão de sub-macroblocos em partições. ....	42
Figura 16 - Codificador Padrão H.266/VVC.....	46
Figura 17 - Particionamento QTMT para CTU de tamanho 128 x128.....	47

## LISTA DE TABELAS

Tabela 1 - Sequência de testes de acordo com o CTC de JVET VVC.....	51
Tabela 2 - Resultado comparativo entre os padrões AVC e VVC para PSNR pelo tempo de codificação.....	52
Tabela 3 - Resultado comparativo entre a taxa de Bitrate pelo tempo de codificação dos padrões AVC e VVC.....	53
Tabela 4 - Resultado comparativo entre os padrões AVC e VVC utilizando resolução 4K.....	54



## LISTA DE ABREVIATURAS E SIGLAS

ALF	<i>Adaptive Loop Filter</i>
AVC	<i>Advanced Video Coding</i>
BITRATE	<i>Número de bits após a codificação</i>
BTH	<i>Binary Tree Horizontal</i>
BTV	<i>Binary Tree Vertical</i>
CABAC	<i>Context-Adaptative Binary Arithmetic Coding</i>
CbCr	<i>Crominância</i>
CC/ALF	<i>Cross-Component Adaptive loop Filtering</i>
CTC	<i>Common Test Condition</i>
CTU	<i>Coding Tree Unit</i>
CU	<i>Coding Units</i>
DBF	<i>Deblocking Filter</i>
DCT	<i>Discrete Cosin Transform</i>
DS	<i>Diamond Search</i>
FPS	<i>Frames per Second</i>
HD	<i>High Definition</i>
HEVC	<i>High Efficiency Video Coding</i>
IPM	<i>Intra prediction modes</i>
ITU	<i>International Telecommunication Union</i>
JVET	<i>Joint Video Experts Team</i>
JVT	<i>Joint Video Team</i>
LCMS	<i>Luma Mapping with Chroma Scaling</i>
MC	<i>Motion Compensation</i>
ME	<i>Motion Estimation</i>
MPEG	<i>Moving Picture Experts Group</i>
MRL	<i>Multiple Reference Lines</i>
MSE	<i>Mean Squared Error</i>
MTT	<i>Multi-Type Tree</i>
Pixel	<i>Picture Elements</i>
PSNR	<i>Peak Signal to-Noise Ratio</i>
QTMT	<i>Quadtree with nested Multi-Type</i>

RD	<i>Rate-Distortion</i>
RGB	<i>Red, Green, Blue</i>
SAO	<i>Sample Adaptive Offset</i>
SNR	<i>Sinal Noise Ruido</i>
SVH	<i>Sistema visual humano</i>
TTH	<i>Ternary Tree Horizontal</i>
TTV	<i>Ternary Tree Vertical</i>
UHD	<i>Ultra High Definition</i>
VVC	<i>Versatile Video Coding</i>
Y	<i>Luminância</i>
ECC	<i>Error Correcting Code</i>
RAM	<i>Random Access Memory</i>

## Sumário

1. Introdução.....	17
1.1. Objetivos .....	19
1.1.1. Objetivo Geral .....	19
1.1.2. Objetivos Específicos.....	19
1.2. Estrutura do Trabalho.....	19
2. Conceitos do Processamento de Imagens Digitais .....	20
2.1. Aquisição de Imagens.....	20
2.2. Processamento de Imagens .....	22
2.3. Representação/ Armazenamento das Imagens .....	24
2.4. Armazenamento .....	26
3. Conceitos de Codificação de Vídeo .....	27
3.1. Redundâncias de dados.....	29
3.2. Etapas de um Codec de vídeo.....	31
4. Evolução dos Codificadores e o Padrão H.264/AVC.....	35
4.1. Avanço dos Padrões de Codificação .....	35
4.2. Conceitos do padrão H.264/AVC.....	37
4.3. Módulo de Estimção de Movimento (ME) .....	39
4.4. Compensação de Movimento (MC) .....	42
5. Padrão H.266/VVC.....	44
5.1. Particionamento de Imagem .....	46
5.2. Ferramentas de Codificação Intra.....	48
5.3. Ferramentas de Codificação Inter .....	48
5.4. Filtros em <i>Loop</i> .....	50
5.5 Resultados Comparativos .....	50
6. Considerações Finais.....	55
REFERÊNCIAS .....	56

## 1. Introdução

Com a constante evolução da tecnologia, tem se percebido um grande crescimento de dispositivos digitais voltados para a visualização e compartilhamento de vídeos. Neste conceito, pode-se destacar diversos produtos, tais como, computadores, smartphones, câmeras digitais, tablets, receptores de televisores HD (do inglês, *High Definition*), dentre outros. A principal causa para o aumento exorbitante da utilização de vídeos ocorre pelos serviços de *Streaming* de vídeo, que é ofertado por grandes empresas, como Amazon, Apple, HBO, Netflix e Youtube. Atualmente, este é o serviço digital mais consumido, chegando a corresponder uma fatia na média de 50% do tráfego da Internet. (SANDVINE, 2020). É importante destacar, que este crescimento acelerado também é uma das consequências da pandemia COVID-19, que, no ano de 2020, a grande maioria das empresas, dos mais variados ramos, tiveram de adaptar a forma de trabalhar, utilizando principalmente aplicações de compartilhamento de vídeo neste período.

Desta forma, tornou-se cada vez mais necessário a disseminação e otimização de técnicas com o objetivo de reduzir o *bitstream* necessário para representar um vídeo, atrelando a qualidade com a redução da taxa de dados sem que haja perda significativa de informação. A partir desta premissa, foram desenvolvidos novos codificadores de vídeos, que fazem parte dos chamamos padrões de codificação de vídeo, que foram desenvolvidos ao longo dos anos.

Dos diversos padrões que foram propostos ao longo dos anos, o padrão H.264/AVC (do inglês, *Advanced Video Coding*), proposto em 2003, foi um grande marco na evolução da codificação de vídeos até então, pois seus resultados em termos de compressão de dados, eficiência na codificação com resolução de alta definição (720p e 1080p), e baixa perda de informação foram um divisor de águas da época, que, em alguns anos, foi consolidado mundialmente como o padrão a ser seguido pela indústria. Porém, com o grande aumento da procura dos consumidores por uma maior qualidade de vídeo, aumentando sua resolução, tornou-se necessário tecnologias que comportassem vídeos de definição ultra alta (UHD, do inglês *Ultra High Definition*), com definições de 4k, 8k e 16k.

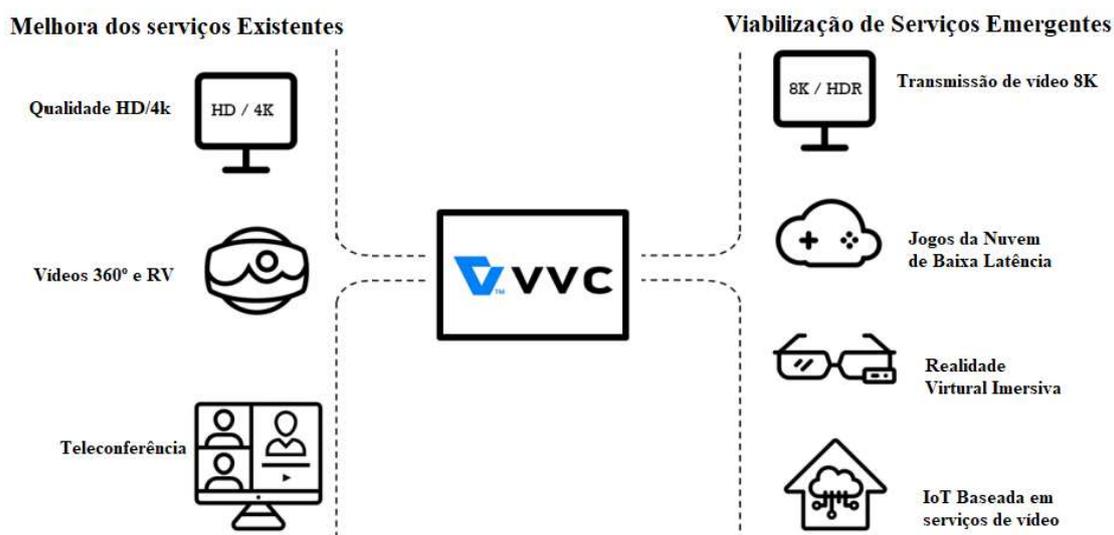
O padrão H.265 também conhecido como HEVC (do inglês, *High Efficiency Video Coding*) é a evolução do padrão H.264/AVC, cujo o objetivo era dobrar as taxas de compressão do padrão AVC, porém sem aumentar muito a complexidade computacional

na codificação de vídeos. Este padrão propôs variadas inovações e otimizações, dentre elas, pode-se destacar a implementação do filtro chamado de Deslocamento Adaptativo de Amostra (do inglês, *Sample Adaptive Offset* – SAO), responsável por reduzir a distorção entre a imagem original e a imagem reconstituída. (SULLIVAN, 2012). Este filtro foi também, posteriormente, implementado pelo padrão H.266/VVC, executando um importante papel em seu codificador. Devido ao pouco tempo de mercado, o padrão HEVC não obteve grande difusão no mercado industrial tecnológico, abrindo espaço para o desenvolvimento de um novo padrão, o VVC.

A partir das demandas por tecnologias com suporte UHD, foi também necessário um grande aumento da complexidade dos algoritmos de codificação e das ferramentas utilizadas neste processo, impossibilitando o uso de tecnologias de padrões passados, sendo necessário propor novos padrões para a codificação de vídeo. Assim, em 2018, surge o padrão de compilação de vídeo versátil, ou VVC, (do inglês, *Versatile Video Coding*), trazendo novas ferramentas e otimizações, com a proposta de tornar viável serviços com altas resolução, com uma eficiência de compressão de até 50% menor que seu antecessor. (CHEN, 2020).

Com sua alta capacidade de compressão, o VVC vem com a ideia de tornar viável diversos serviços emergentes nos próximos anos, como mostra a Figura 1.

Figura 1 - Potenciais melhoras e novos serviços do padrão VVC.



Fonte: Modificado de HAMIDOUICHE, 2021.

Neste contexto, o presente trabalho aborda os conceitos da compressão de vídeo, apresentando de uma forma geral o processo desde a representação da imagem à codificação de vídeos. A proposta consiste em apresentar um estudo amplo dos conceitos dos codificadores dos padrões H.264/AVC e H.266/VVC, discutindo suas características e trazendo suas principais características, evoluções e otimizações de acordo com suas propostas.

## **1.1. Objetivos**

### **1.1.1. Objetivo Geral**

O objetivo geral deste trabalho é apresentar um estudo dos principais conceitos abordados nos padrões de codificação de vídeo H.264/AVC e H.266/VVC, abordando seus respectivos codificadores e decodificadores (CODECs) e trazendo uma comparação entre seus parâmetros de eficiência.

### **1.1.2. Objetivos Específicos**

- Abordar os conceitos básicos do processamento de imagens digitais;
- analisar os conceitos da codificação de vídeos;
- levantar as características da evolução dos padrões de codificadores de vídeo ao longo dos anos;
- apresentar os módulos de codificação dos padrões H.264 e H.266.

## **1.2. Estrutura do Trabalho**

Este trabalho de conclusão de curso está organizado em 05 (cinco) capítulos, sendo eles:

- Capítulo 1 – Contextualização do tema de forma introdutória, a definição dos objetivos e a estrutura do trabalho;
- Capítulo 2 – Apresenta os conceitos do processamento de imagens;
- Capítulo 3 – Descrição dos princípios gerais da codificação de vídeo;
- Capítulo 4 – Apresenta o estudo sobre as estruturas do padrão H.264/AVC;

- Capítulo 5 – Apresenta o estudo sobre as estruturas do padrão H.266/VCC.

## **2. Conceitos do Processamento de Imagens Digitais**

Neste capítulo são apresentados os conceitos do processamento de imagens que auxiliarão o entendimento do presente trabalho, passando pela definição sucinta de quatro etapas principais no processo de captura de imagens. O objetivo deste capítulo é introduzir ao leitor conceitos simples, porém essenciais para adentrar ao assunto de codificação de vídeo.

Do processamento de sinais deriva-se o conceito de Processamento de Imagens. As imagens, assim como os sinais, carregam uma determinada informação. O processamento de imagens nada mais é do que a extração dessas informações para um objetivo específico. O sistema visual humano (SVH) é o melhor e mais complexo processador de imagens existente atualmente, logo, processar uma imagem digital como feito pelo SVH é uma tarefa muito complexa. Estes aspectos fazem com que o processamento de imagens seja um tema com enorme dependência ao sistema e finalidade para o qual está sendo usado, não existindo apenas uma solução abrangente para todos os casos (BETETTO, 2005). Portanto, até o momento, não existem sistemas de análise de imagens que funcionem para todo e qualquer processamento de Imagens (ALBUQUERQUE &ALBUQUERQUE, 2009).

As imagens, do ponto visual ótico, podem ser definidas pelo conjunto de pontos que convergem em um plano bidimensional, ou do senso comum, um objeto que permite a troca de informações digitais.

O sistema para o processamento de imagens digitais pode ser classificado nas seguintes etapas (BETETTO, 2005):

- Aquisição das Imagens;
- processamento ou Digitalização;
- representação;
- armazenamento.

### **2.1. Aquisição de Imagens**

Os sistemas construídos com a função de captura das imagens são chamados sistemas de imageamento. Atualmente, existem diversos sistemas com esta finalidade,

mas pode-se destacar os mais comuns como sendo sistemas de máquina e filme fotográfico, que são capazes de registrar as ondas eletromagnéticas presentes numa cena real que são perceptíveis ao SVH. (ZAGHETTO, 2009).

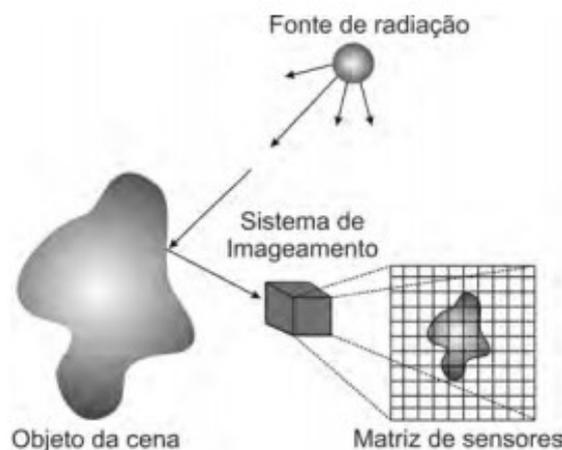
O plano bidimensional no qual a imagem pode ser definida, é representado por suas coordenadas espaciais  $X$  e  $Y$ , que formam uma função bidimensional  $f(x, y)$ , onde sua amplitude é a intensidade da imagem naquele ponto.

As imagens geradas a partir de sistemas físicos, como por exemplo, filmes fotográficos, caracterizam a função  $f(x, y)$  como sendo a luz incidente na cena de captação,  $i(x, y)$ , e da porção de luz refletida no objeto de análise,  $r(x, y)$ . Logo, pode-se destacar a equação 2.1.

$$f(x, y) = i(x, y)r(x, y) \quad (2.1)$$

Onde  $i(x, y)$  e  $r(x, y)$  são denominados, respectivamente: Iluminância e reflectância. Pode-se expressar o processo de aquisição da imagem de acordo com a Figura 2.

Figura 2 - Sistema de Aquisição de Imagem.



Fonte: ZAGHETTO, 2009.

Em que o objeto a ser capturado é incidido por uma fonte de radiação luminosa, onde um sistema de imageamento registram as ondas eletromagnéticas presentes na reflexão da fonte luminosa sobre o objeto da cena, criando uma matriz de sensores.

Para se representar uma imagem e, conseqüentemente um vídeo, existem diversas formas de se expressar as cores no modelo digital. A definição de tal forma de expressão é primordial para se conseguir uma boa eficiência de codificação da imagem.

Pode-se dividir as imagens em dois segmentos: imagens coloridas e imagens em escala de cinza.

As imagens em escala de cinza são aquelas que, quando codificadas, não possuem informação de cores, em termos técnicos são chamadas monocromáticas. Para essas imagens, variam do preto para o branco, é neste intervalo que são expostos os tons de cinza, que sugerem a luminância, ou seja, a intensidade luminosa da imagem. (GONZALES, 2003).

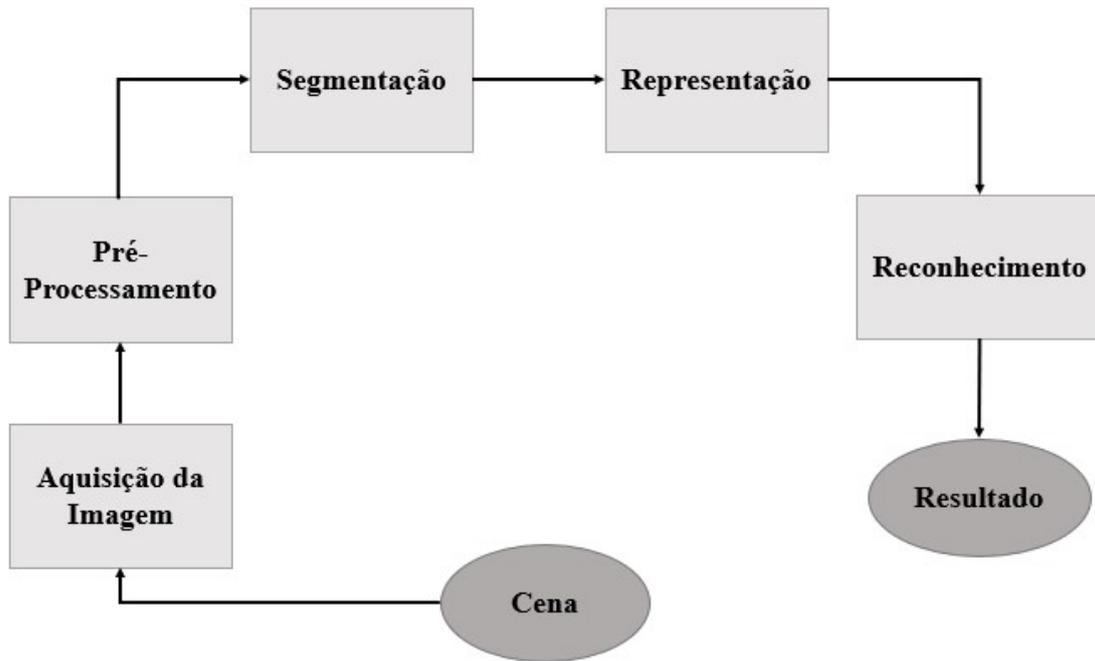
Para as imagens coloridas, existem diversos espaços de cores que podem representá-las. Como, por exemplo, o espaço de cores RGB (do inglês, *Red, Green, Blue*). A imagem é definida a partir de três matrizes distintas, responsáveis por registrar as intensidades primárias, vermelha, verde e azul, que são refletidas por uma cena. (WYSZECKI, 2000)

O RGB, por sua vez, possui níveis de iluminância espalhados igualmente entre suas três matrizes, ou seja, estão correlacionadas e sabe-se, segundo GONZALES, que o SVH é mais sensível ao brilho (Luminância) do que a própria cor (Crominância). Logo, de acordo com essas afirmações, não é o espaço de cores ideal se o objetivo é compressão de imagens e vídeo. (GONZALES, 2003). O espaço denominado YCbCr, trabalha com os componentes luminância e crominância de forma separada, ou seja, sua cor (Crominância - CbCr) está inteiramente separada de seu brilho (Luminância - Y). Sendo assim, se tratando de compressão de imagens e vídeos pode-se “aumentar a eficiência de codificação através da redução de taxa de amostragem dos componentes de crominância em relação aos componentes de luminância”, conforme Agostini (2007, pág. 30), sem prejudicar a qualidade final da imagem para o sistema visual humano, pois será imperceptível ao SVH esta redução, esta técnica é chamada de subamostragem de cores. (RICHARDSON, 2003).

## **2.2. Processamento de Imagens**

De acordo com a Figura 3, o processamento de Imagens é composto de maneira geral pelos seguintes blocos que visam maximizar o reconhecimento dos padrões que a imagem carrega em toda sua informação existente.

Figura 3 - Níveis Básicos do Processamento de Imagens.



Fonte: Autoria própria.

Através de uma cena qualquer, pode-se realizar a aquisição de uma imagem, que por sua vez, pode ser capturada através de diversos dispositivos capazes de digitalizá-la. O pré-processamento busca um certo nível de melhoramento da imagem, em termos de realce do contraste, atenuação de ruídos e etc. (GONZALES, 2002).

A segmentação por sua vez, constitui-se das subdivisões da imagem em fragmentos ou em objetos que dão forma à imagem, facilitando o processo de reconhecimento e análise dessas porções (GONZALLES, 2002). Esta etapa tem como objetivo capturar regiões da imagem original com características mais relevantes para um determinado fim, separando os fragmentos que possuem características em comum e que resultam numa redundância que não contribui com informações relevantes para o bloco de representação da imagem (AGOSTINI, 2007).

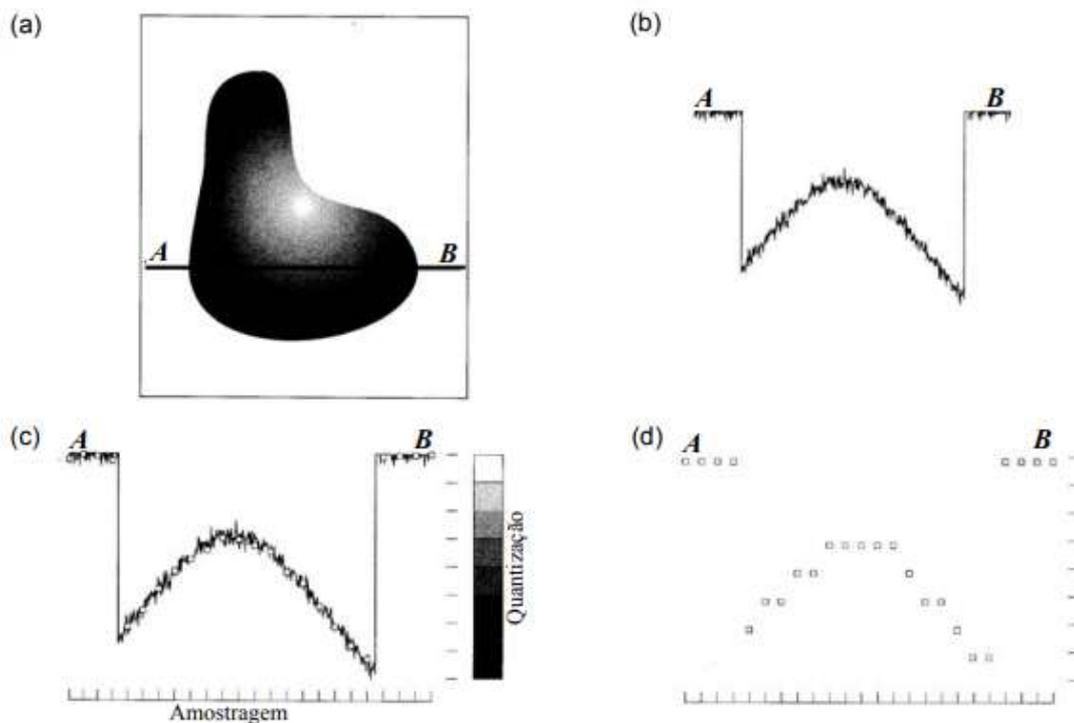
A representação discretiza a imagem em termos de amostragem e quantização da mesma em um espaço finito de tempo, seguido pelo reconhecimento que restitui o conjunto de fragmentos, atribuindo seu significado ao resultado da imagem.

### 2.3. Representação/ Armazenamento das Imagens

O processo de representação da Imagem consiste na discretização das coordenadas espaciais,  $f(x, y)$ , presentes na equação 2.1 e sua intensidade luminosa.

Esta discretização é feita através da amostragem da função  $f(x, y)$ , onde é estabelecido matrizes de cores com pontos igualmente espaçados entre si. O processo de quantização define o conjunto de níveis de cinza desta matriz, ou seja, a resolução em níveis de cinza da imagem (ZAGUETO, 2006).

Figura 4 - a) Imagem Contínua, (b). Uma Linha Escaneada de A para B na Imagem Contínua, (c) Amostragem e Quantização, (d) Linha Digital Escaneada.



Fonte: GONZALES, 2002.

A Figura 4 retrata uma imagem contínua ao longo de uma linha do segmento AB. A Figura 4(b) plota os valores da amplitude em níveis de cinza da Figura 4(a), ou seja, processo de quantização da imagem. A Figura 4(c) representa o processo de amostragem da imagem para o segmento AB e seus níveis de quantização ao lado direito. O resultado digital deste processo é exibido na Figura 4(d). (GONZALES, 2002).

As imagens são, basicamente, compostas por elementos normalmente em formato retangular, denominados pixels (do inglês, *Picture Elements*). Logo, pode-se definir uma imagem como uma matriz de pixels,  $F(m, n)$ , onde sua amplitude  $K$  é caracterizado como

a quantização em níveis de cinza da imagem, M e N sua posição dentro da matriz, linha e coluna, respectivamente.

Então, é possível expressar uma matriz de pixels da seguinte forma:

$$F(m, n) = \begin{pmatrix} F(0,1) & \cdots & F(0, N - 1) \\ \vdots & \ddots & \vdots \\ F(M - 1, 0) & \cdots & F(M - 1, N - 1) \end{pmatrix} \quad (2.2)$$

Onde, segundo GONZALLES, a matriz F (m, n) não armazena os valores físicos de luminosidade, mas os índices numéricos intrínsecos em cada pixel presente na matriz. Assim, essa imagem possui K níveis de escala de cinza dentro de um intervalo [0, K-1], para os níveis preto e branco, respectivamente. (GONZALLES, 2002).

Dessa forma, os valores de K se dão através da Equação 2.3.

$$K = 2^b \quad (2.3)$$

Onde b é o número de bits que representa cada pixel presente na matriz. Para exemplificar, se b = 4 bits, vemos que  $K = 2^4 = 16$  níveis de cinza. Então, os valores da matriz F (m, n) estão entre 0 e 16.

Finalmente, pode ser calculado o número de bits necessários para o armazenamento de uma imagem de acordo com a matriz M x N e com K níveis de cinza de acordo com a Equação 2.3, sendo obtida através da Equação 2.4.

$$N_{bits} = M \times N \times \log_2(K) \quad (2.4)$$

Seguindo o exemplo anterior e considerando uma matriz composta por 4 linhas e 4 colunas, têm-se  $N_{bits} = 4 \times 4 \times \log_2(16) = 64$  bits.

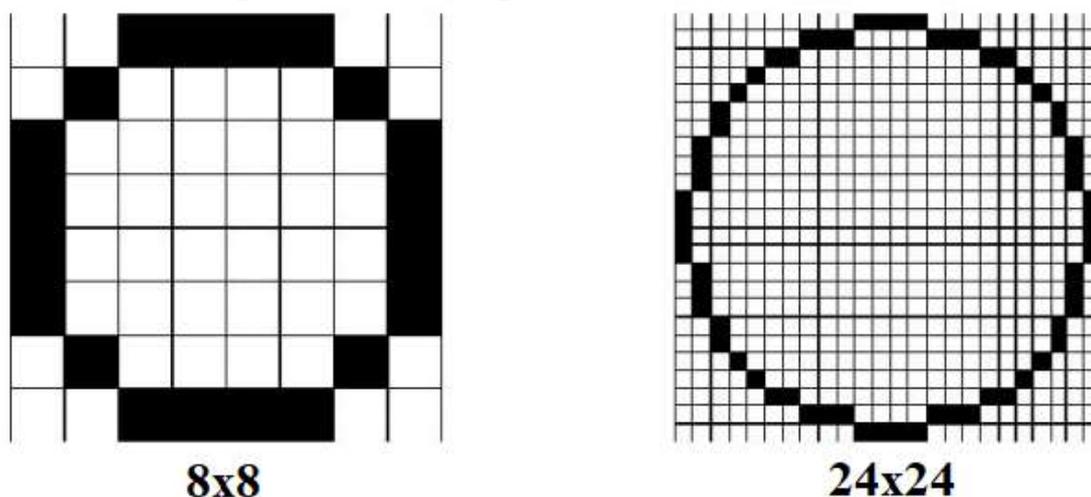
Desta forma, estabelece-se um ponto de partida para a entrada das técnicas de compressão, que são extremamente utilizadas nos formatos atuais. Este trabalho abordará a evolução das técnicas de compressão do novo padrão H.266 e do padrão H.264.

## 2.4. Armazenamento

O armazenamento das imagens é de grande interesse, pois utiliza-se espaço de memória para guardá-las. Porém, sabe-se que quanto melhor for a qualidade visual dessas imagens, maior será a quantidade de pixels presentes na mesma, então o volume de dados a serem transmitidos aumentam e, conseqüentemente, exigem maior capacidade de armazenamentos, influenciando em um maior número de bits utilizados para a codificação da imagem, que implica numa maior requisição de espaço de memória (SCURI, 2002).

A Figura 5 exemplifica imagens com o mesmo tamanho, mas com resoluções diferentes de acordo com sua quantidade de pixels que cada uma carrega.

Figura 5 - Mesma Imagem com Resoluções Diferentes.



Fonte: Adaptado de SCURI, 2002.

Como mencionado anteriormente, as técnicas de compressão são muito úteis para que se possa transmitir e armazenar imagens em alta resolução. Tais técnicas podem ser divididas em duas vertentes, compressão com ou sem perda de informação. De maneira simplória, ambas utilizam como base a redução da quantidade de dados redundantes nas informações das imagens ou vídeo.

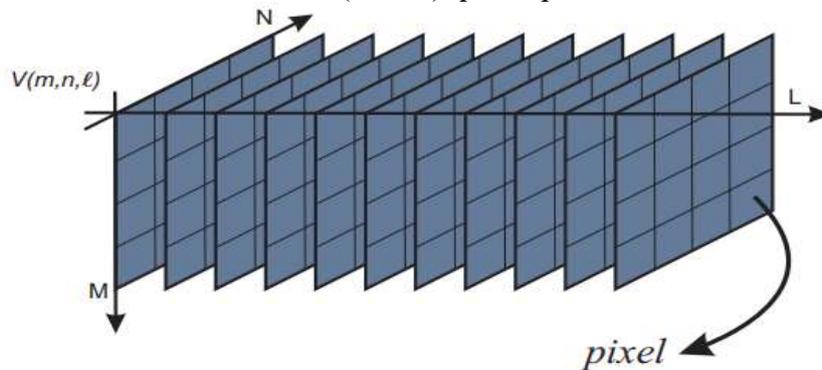
### 3. Conceitos de Codificação de Vídeo

Este capítulo abordará, de forma sucinta, os conceitos básicos da codificação de vídeo, levantando os tipos de redundâncias presentes na codificação de vídeo e as ferramentas necessárias para tratá-la. Além disso, é apresentado o modelo de um codificador elementar e suas etapas.

Um vídeo é uma sequência de cenas em um espaço de tempo. Com os conceitos abordados no capítulo 2, após o processamento de uma imagem, um vídeo é capaz de ser representado.

Durante um intervalo de tempo  $\Delta t$ , cada cena capturada é chamada de quadro (do Inglês, *Frame*). Logo, a partir da matriz de pixel bidimensional (Eq. 2.2) de uma imagem, considerando cada uma como um quadro de um vídeo natural, é possível representar um vídeo digital como uma matriz tridimensional de pixels,  $V(m, n, l)$ , com  $M$  linhas,  $N$  colunas e  $L$  quadros, onde cada pixel caracteriza  $K$  níveis de cinza conforme Equação 2.3 (ZAGUETO, 2006). A Figura 6 ilustra esta matriz.

Figura 6 - Matriz Tridimensional  $V(m, n, l)$  que Representa um Vídeo Digitalizado.

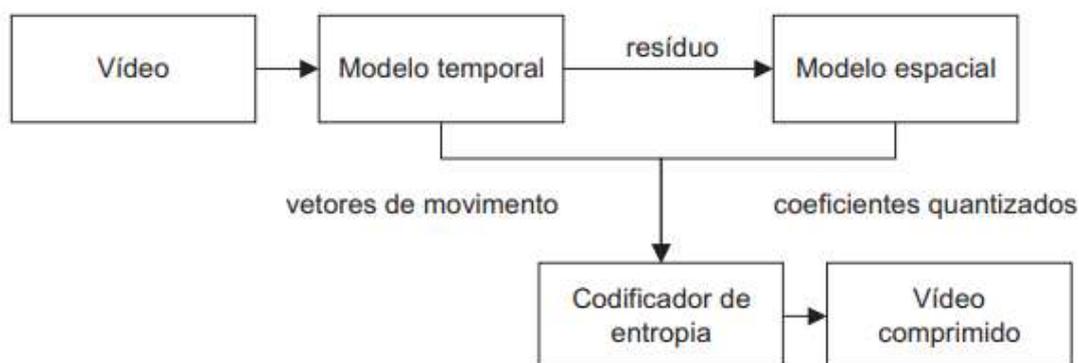


Fonte: ZAGUETO, 2006.

Para se ter ideia do tamanho de um vídeo bruto, supõe-se que um vídeo com duração de noventa minutos possui uma taxa de 60 FPS (do inglês, *Frames per Second*), onde cada quadro apresenta uma matriz com  $M = 234$  linhas e  $N = 318$  colunas. Neste caso, o vídeo em exemplo teria um número de quadros total  $L = 1,5 * 3600 * 60 = 324000$  quadros, onde, cada quadro teria  $234 * 318 = 74412$  pixels. Se cada um desses pixels for codificado com 8 bits, o vídeo completo teria aproximadamente 24,1 GB de tamanho.

Um codificador é um conjunto de blocos com o intuito de realizar a compressão de, neste caso, vídeos. Já o decodificador é responsável por decodificar esses dados comprimidos com a menor taxa de erro possível. Assim, pode-se denominar um dispositivo de transmissão/recepção de vídeo como CODEC. O diagrama de blocos abaixo apresenta os blocos que envolvem um CODEC básico contendo: um modelo temporal, modelo espacial e um codificador por entropia.

Figura 7 - Diagrama de Blocos de um CODEC Básico.



Fonte: ZAGUETTO, 2006.

Inicialmente, uma das estratégias para se codificar e comprimir um vídeo seria enviar apenas as diferenças entre os frames por um segundo, esta técnica aborda a redundância temporal do sinal de entrada do vídeo, denominada codificação interquadros (SAYOOD, 2000). Esta codificação permite reduzir consideravelmente o tamanho de um conjunto de frames, visto que apenas pequenas partes de cada um será enviada. Porém, para realizar esta codificação é necessário ter, ao menos, dois quadros, visto que não se pode comparar um quadro com ele mesmo. Logo, é necessário utilizar técnicas que analisem também a redundância no interior de um mesmo quadro, esta codificação é denominada intraquadros, e explora as redundâncias espaciais alocadas neste quadro. Posteriormente, será abordado com mais detalhes os tipos de redundâncias e como elas ocorrem.

Um Codec, usualmente, como descrito na Figura 7 é composto pelas técnicas de codificação interquadros, que abordam a modelagem temporal, e intraquadros que analisam a modelagem espacial (SEQUEIRA, 2019). O resíduo gerado do modelo temporal é justamente a diferença entre o quadro anterior (Quadro Predito) e o quadro atual (Quadro de Referência), podendo nomeá-lo como quadro residual. Em seguida,

buscam-se nos quadros de referência blocos preditores que melhor predizem o bloco atual, realizando a estimação de movimento de um bloco dentro de quadro atual, gerando vetores de movimento, que relatam o deslocamento entre os blocos atuais. (ZAGUETTO, 2006).

Por sua vez, o quadro residual gerado é tratado no modelo espacial, reduzindo as redundâncias entre as próprias amostras residuais vizinhas, através da transformação e quantização. Por fim, os coeficientes quantizados são codificados por entropia, junto com os vetores de movimento.

### 3.1. Redundâncias de dados

De acordo com Agostini (2007), “a codificação de vídeos busca diminuir a quantidade de dados considerados redundantes na representação computacional das informações das imagens ou vídeo”. Isto é, através da codificação são excluídos todos os dados que não contribuem de maneira significativa na mensagem da informação, filtrando somente o que irá influenciar diretamente na imagem ou vídeo a ser recebido.

A partir disto, estudam-se algoritmos de compressão de dados que buscam maximizar a identificação e o tratamento dessas redundâncias presentes na codificação de um vídeo. Para codificação/compressão de vídeos pode-se destacar quatro diferentes tipos de redundância: espacial, temporal, entrópica e também a redundância psicovisual (MONTEIRO, 2013).

- **Redundância Espacial** - Conhecida também como redundância intraquadros ou Interpixel. Com ela, pode-se prever o valor de um pixel pelo valor de seus pixels vizinhos através da correlação espacial que está ligada ao posicionamento geométrico distribuído em um mesmo quadro. (GONZALES, 2002). Isto é, pode-se assumir que, quanto maior for a resolução deste quadro, mais visual é esta redundância, pois quanto maior o número de pixels presentes em um quadro, maior será a tendência de um pixel vizinho possuir valores similares. Como mencionado a partir da Figura 6, esta redundância pode ser amenizada através da codificação intraquadros.

- **Redundância Temporal** - Conhecida também como redundância interquadros. Diferentemente da espacial, esta redundância ocorre pela correlação de um conjunto de quadros próximos em um vídeo. Neste caso, muitos blocos de pixels não mudam o valor com o passar dos quadros, sejam em um fundo que não foi alterado ou um objeto fixo em uma cena de diálogo. Outrora, estes blocos podem apresentar pequenas variações, por exemplo, de iluminação, mas que interferem muito pouco ou nada na qualidade do vídeo. Também se têm a situação em que um bloco de pixels apenas se desloque entre os quadros, tendo alteração nula ou quase nula no seu valor. (AGOSTINI, 2007). Esta redundância é onde, atualmente, se ganha altas taxas de compressão, devido as diferentes situações presentes nos vídeos onde ela ocorre.
  
- **Redundância Entrópica** – Esta redundância está relacionada com a representação dos dados a serem codificados. Basicamente, quanto maior for a probabilidade de ocorrência de um símbolo, menor a quantidade de informações novas este símbolo transmite (AGOSTINI, 2007). Ou seja, nesta redundância, os codificadores buscam representar as informações mais frequentes de um vídeo através de um único símbolo codificado, gerando uma menor taxa de bits para representar maior informação. Neste caso, não há perda de informação no processo de compressão.
  
- **Redundância Psicovisual** - Como o próprio nome sugere, trata-se de uma redundância focada no sistema visual humano (SVH). Em relação ao SVH, algumas informações passam despercebidas, pois o olho humano não é sensível a todos os níveis de informação de cores e luminosidade (MONTEIRO, 2003). Através deste fato, os codificadores de vídeo buscam reduzir a amostragem dos componentes de crominância, como abordado anteriormente na seção 2.1, enviando apenas informações essenciais para o SVH. Neste caso, a redução desta redundância leva a uma compressão com perdas, visto que essas componentes são eliminadas na codificação.

Vários algoritmos são desenvolvidos com o intuito de otimizar a redução destas redundâncias, onde têm-se uma maior compressão de dados. Este trabalho visa evidenciar algoritmos de comparação que analisam tais redundâncias no padrão H.264 ao H.266.

### 3.2. Etapas de um Codec de vídeo

A figura 7 representa o fluxo de um Codec básico, utilizados pelos padrões atuais de codificação. O processo se inicia a partir da predição, intra e interquadros, onde são tratadas as redundâncias espaciais e temporais, respectivamente, que foram descritas anteriormente.

Existem diversos modos de predição inter, que são baseadas nas técnicas de Estimção de movimento e Compensação de movimento para encontrar as redundâncias temporais, iniciando com a escolha dos quadros de referência. Existem vários algoritmos de buscas que atendem ao objetivo, pode-se destacar o algoritmo *Diamond Search* (DS), que analisa todos os blocos candidatos de uma determinada área através de padrão em forma de diamante, possuindo como vantagem sua baixa complexidade computacional (PORTO, 2012). Após essa etapa, as comparações entre os blocos são realizadas através da similaridade entre eles, utilizando, por exemplo, o critério do erro quadrático médio (do inglês, *Mean Squared Error* – MSE). (KUHN, 1999).

Já para a predição intraquadros, vale destacar entre os modos mais comuns, o modo DC, que gera um bloco predito onde a média das amostras de referência são base para todas as amostras preditas. Também é válido mencionar o modo Direcional, que é gerado através da extrapolação das amostras dos blocos adjacentes a partir do ângulo que compõe o bloco predito (WIEGAND et al, 2003).

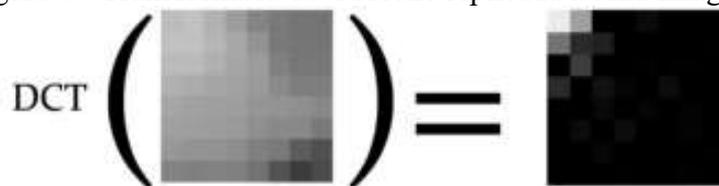
Ambos esses blocos, reduzem exponencialmente o volume de dados a ser codificados, e abrem espaço para diversas oportunidades de pesquisa em busca de otimização dos mesmos.

Após a etapa de predição, são gerados os quadros resíduos, que são a diferença entre o quadro reestruturado através das predições e o quadro original. Estes resíduos então chegam à etapa seguinte, Transformada (T), onde tem o objetivo de converter os dados residuais no domínio do tempo para o domínio da frequência.

Neste bloco, os tipos de transformadas são variáveis de acordo com o padrão de codificação a ser usado. Porém, destaca-se por seu uso comum, a transformada do cosseno (do inglês, *Discrete Cosin Transform* – DCT).

A DCT é muito utilizada nos padrões de compressão de vídeo pela sua alta capacidade de eficiência na compactação da energia de um sinal e sua decorrelação. Onde, a compactação da energia é o nome dado à técnica de uma transformada alcançar a compactação da energia de um sinal na menor quantidade possível de coeficientes. Por sua vez, a decorrelação é outra característica da transformada, em que converte um determinado conjunto de dados que são muito correlacionados, ou seja, possuem uma relação de correspondência, em um outro conjunto que possuem maior independência (WU, 2009).

Figura 8 - Transformada do Cosseno Aplicada a uma Imagem.



Fonte: ROSA, 2010.

A figura 8 relata a aplicação da DCT a uma imagem. A transformada se baseia no fato de para cada bloco a ser usado, existe uma matriz DCT fixa (WU, 2009). No exemplo, as amostras da matriz original são de 8 bits, logo, sua transformada implica em analisar os 64 pixels dessa imagem, realizar a DCT, obtendo 64 novos valores, os coeficientes DCT. Após a aplicação da DCT, a matriz resultante possui o primeiro coeficiente (canto superior esquerdo) de maior valor e seus últimos coeficientes (mais próximos ao canto inferior direito) com os menores valores. Sabe-se da seção 2.3, que o valor 0 corresponde ao preto, então os valores muito baixos da matriz resultante são desprezados, sem que a imagem sofra perdas significativas de qualidade, este conceito explora a redundância psicovisual, sendo a maior razão pela qual essa transformada é usada.

Em seguida, têm-se a Quantização direta (Q), que explora também as características de redundância psicovisual, atenuando os coeficientes que passaram pela transformada com maior e menor intensidade de acordo com sua frequência (AGOSTINI, 2007), resultando em matrizes esparsas e potencializando o próximo bloco, a codificação por Entropia.

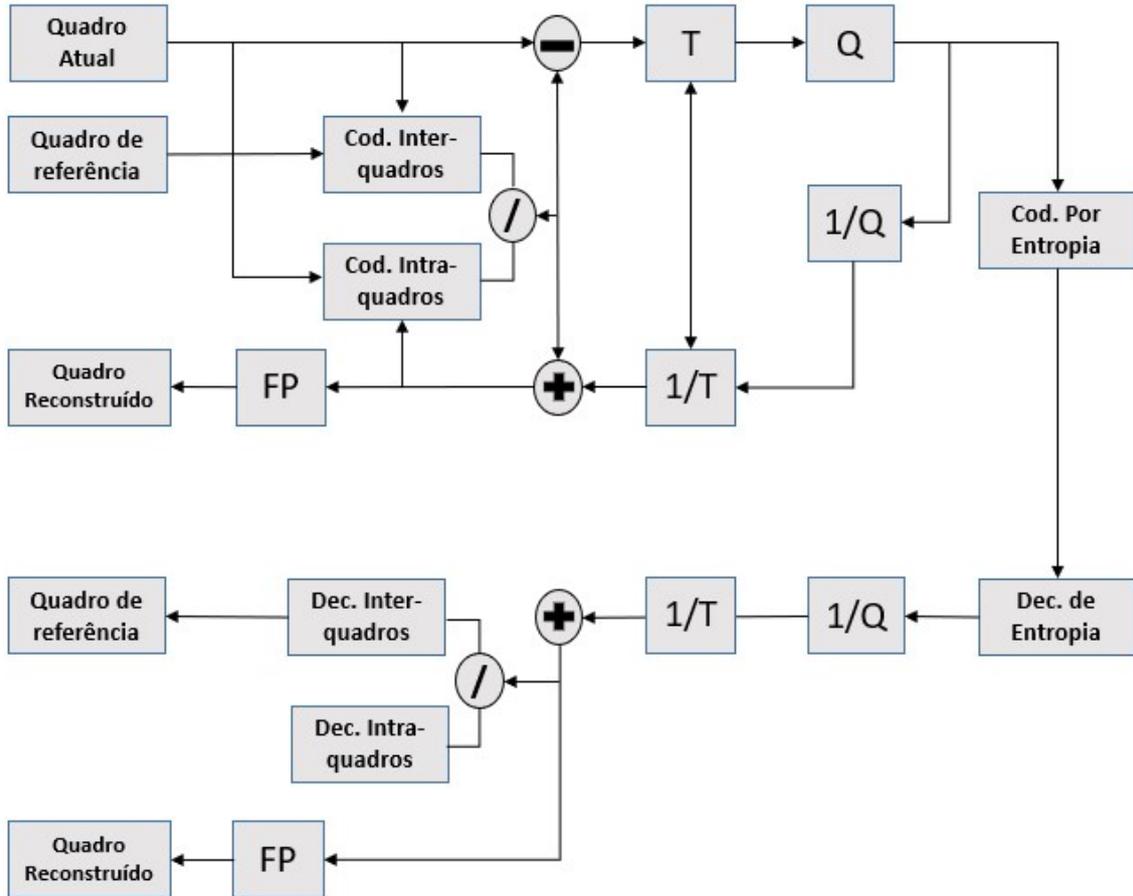
Por sua vez, a codificação entrópica é aplicada nas amostras residuais após serem transformadas para o domínio da frequência e quantizadas, gerando um fluxo de bits a ser decodificado. A transformada inversa e quantização inversa são agregadas ao codificador com o objetivo de atualizar o quadro de referência, pois sem os mesmos, quando um próximo quadro é processado, o quadro atual que ainda não foi codificado passa a ser o de referência, fazendo com que o decodificador não receba o original, apenas o codificado. Dessa forma, o quadro de referência do decodificador passa a ser o reconstruído, mas no processo geram-se perdas. Assim, o codificador e o decodificador utilizariam quadros diferentes como referência, o que não é aplicável para o processo (GOEBEL, 2019).

Para reconstruir o próximo quadro de referência, é aplicado o processo de QI e TI, onde os resíduos gerados são somados às amostras geradas pelas previsões inter e intraquadro, gerando uma versão reconstituída do quadro codificado, usando-o como referência para a próxima previsão e assim sucessivamente. Por fim, aplica-se um filtro de modo a suavizar os itens que são inseridos no processo de subdivisão do vídeo.

O processo de decodificação, de forma sucinta, desenvolve o processo inverso do codificador. A decodificação por entropia refaz as amostras codificadas, que por não resultar perdas na informação, resulta no mesmo desfecho após a quantização do codificador (AGOSTINI, 2007). Em seguida, as amostras passam pelo Quantizador Inverso ( $1/Q$ ) e Transformada Inversa ( $1/T$ ), formando o bloco residual, como pode ser visto na Figura 9.

Para se gerar os blocos reconstruídos, a decodificação interquadros utiliza o quadro de referência para identificar os blocos, vale salientar que esta referência deve ser a mesma referência usada para a codificação, caso contrário não será possível reconstruir o bloco predito. Então, o resíduo resultando da  $1/T$  é somado aos blocos extraídos do quadro de referência, gerando os blocos reconstruídos. Enfim, estes blocos são armazenados para a execução do vídeo decodificado e também utilizados na decodificação intraquadros para serem usados como referência ao próximo quadro que será decodificado

Figura 9 - Modelo de um Codec de Vídeo



Fonte: Adaptado de AGOSTINI, 2007.

É importante destacar que o diagrama apresentado na Figura 9 é o modelo genérico de um Codec de vídeo, que foi a base para os estudos e otimizações de Codec's seguintes.

## 4. Evolução dos Codificadores e o Padrão H.264/AVC

Este capítulo é composto por um resumo sobre o padrão de compressão de vídeo H.264/AVC. Primeiramente, é abordada uma rápida introdução ao histórico evolutivo dos padrões de compressão e, por fim, aos módulos que fazem parte ao Codec H.264/AVC, com foco na compensação e estimação de movimento.

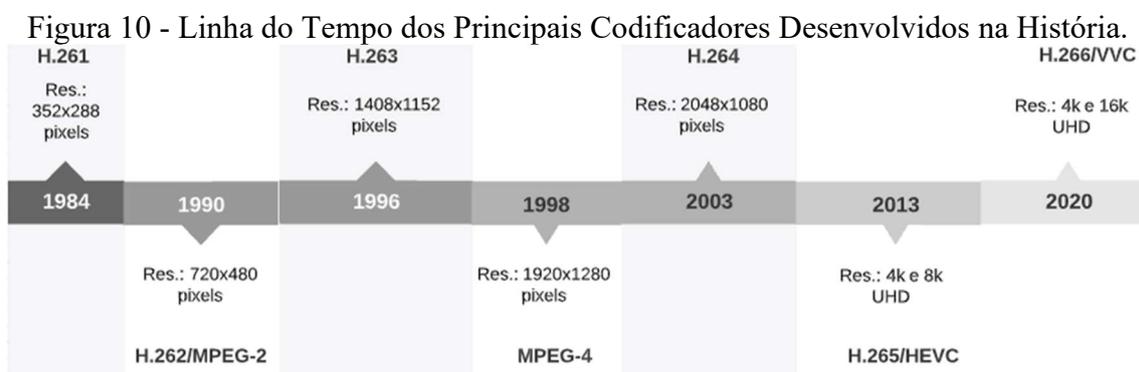
### 4.1. Avanço dos Padrões de Codificação

Obviamente, com o passar dos anos os padrões de codificação obtiveram uma evolução gigantesca. Esse tópico visa trazer um breve resumo dos padrões de codificações que mais marcaram a história da codificação de vídeo até os dias de hoje.

No meio digital, as organizações ITU (do inglês, *International Telecommunication Union*) e MPEG (do inglês, *Moving Picture Experts Group*) especificaram diferentes arquiteturas e padrões de tecnologias relacionados as telecomunicações.

O setor de Normatização das Telecomunicações (ITU-T) é um setor pertencente a ITU, fundado em 1947 como um comitê especializado da ONU. Atualmente é uma organização intergovernamental que é referência para o desenvolvimento de soluções tecnológicas de redes e telecomunicações (ITU-T, 2021).

A Figura 10 mostra o ano de lançamento dos principais padrões de codificação ao longo dos anos. Se for analisado apenas a resolução das imagens, é notório a evolução dos padrões, visto que o primeiro padrão, H.261, resultava numa qualidade de 352x288 pixels e hoje, o padrão H.266, fornece uma qualidade de imagem em 8k e 16k pixels em UHD.



Fonte: Autoria própria.

A primeira geração de codificadores de vídeo é denominada H.261, registrada em 1984, onde foi o primeiro lançamento da organização ITU-T, que visava estabelecer um padrão para o conceito de teleconferências. Tempos depois, o grupo MPEG lançava o padrão MPEG-1, o mesmo foi baseado no padrão H.261, visando além das teleconferências, o armazenamento de vídeo em um disco. Estes foram pioneiros em utilizar o modelo simplificado do CODEC apresentado no Figura 9. Porém, como modelos primários, não possuem boa flexibilidade e escalabilidade do sistema,

A segunda geração foi iniciada em 1990, o padrão H.262/MPEG-2, onde foi criado a partir da parceria entre as organizações MPEG e ITU-T, este padrão é ainda hoje, utilizado para diversas aplicações, como, por exemplo, em discos de DVD. Nesta atualização, os conceitos de escalabilidade temporal, espacial, SNR (do inglês, *Signal Noise Ruido*) e híbrida foram previstos nesta norma (HUSEMAN et. al, 2006). Cujo objetivo era alcançar uma taxa de vídeo de 1,5 Mbps à 15 Mbps para sinais de televisão padrão e taxas de 15Mbps à 30 Mbps para sinais de televisão em HD.

A terceira geração foi lançado em 1996, o padrão H.263, visando trabalhar com quadros de maior definição e menor taxa de bits para videoconferências. Esta otimização não foi tão difundida na comunidade em virtude do ganho de desempenho ser considerado baixo e o modelo anterior já ter se difundido na tecnologia da época. Desde seu lançamento, foram desenvolvidas diversas extensões a fim de otimização da escalabilidade, do qual a primeira versão não contemplava.

A especificação MPEG-4 se estabeleceu como a quarta geração dos padrões de vídeo. Criado ao fim dos anos 90, este padrão buscava com transmissão de vídeos de alta definição com maior taxa de compressão de dados, e também já vislumbrava o início de mercado de telefonia móvel. São várias as vantagens desse modelo, podendo destacar que este padrão utilizou, na época, implementação inovadora, onde possibilitava a análise simultânea de informações em dois domínios diferentes, utilizando, de maneira natural, a escalabilidade híbrida (CHEN, PENG, 2002). Além disso, esta versão foi a primeira a adotar o método DCT para compressão de dados, adotando um mecanismo de planos de bit através dos coeficientes DCT para a construção de sua codificação.

A norma H.264 foi lançada no ano de 2003, este modelo é amplamente utilizado ainda hoje, devido a sua alta compatibilidade com diferentes dispositivos e algoritmos. Criado pelo grupo JVT (do inglês, *Joint Video Team*) o padrão H.264 pode ser considerado um divisor de águas para o mundo digital, por sua grande compatibilidade a vídeos de alta e baixa resolução/taxa de bits. A partir desse modelo, foram criadas

diversas extensões do mesmo para um determinado objetivo. As gerações sucessoras, os padrões H.265 e H.266 tiveram suas primeiras versões publicadas no ano de 2013 e 2018, respectivamente. O padrão H.265/HEVC foi elaborado pelo grupo JCT-VC, com várias ferramentas e características que os anteriores não tinham, além de uma menor taxa de bits (BENDER, 2020). O padrão H.266 ou VVC, teve sua norma finalizada em 2020, portanto ainda é uma tecnologia que está em fase de implementação.

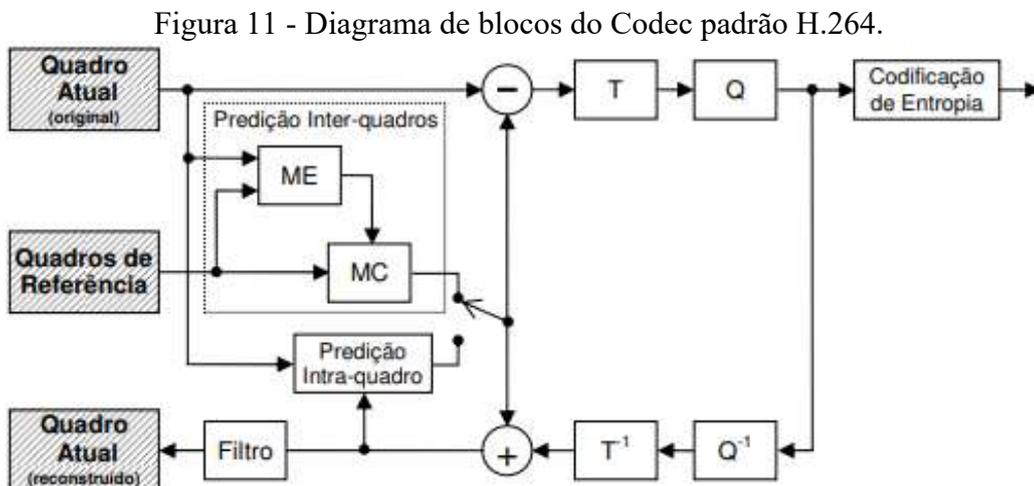
Os padrões H.264 e H.266 são modelos que estão ou serão muito utilizados para compressão de vídeo atuais, então suas estruturas básicas devem ser apresentadas de maneira sólida. Este texto abordará estas versões mencionadas com maiores detalhes nas próximas páginas.

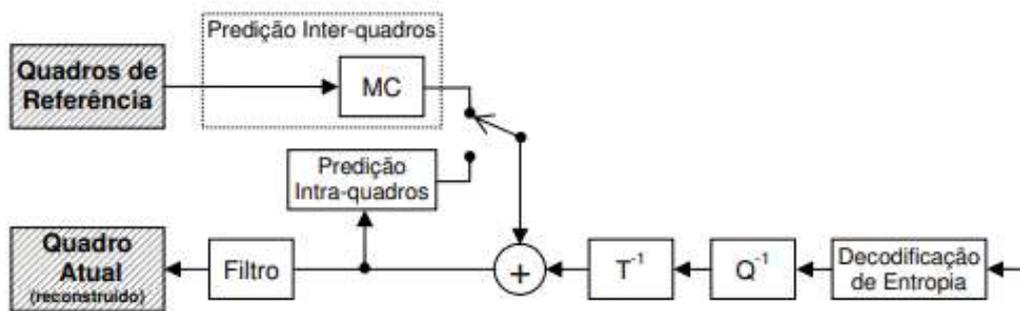
#### 4.2. Conceitos do padrão H.264/AVC

O padrão H.264 oferece elevada eficiência de compressão à classe de sinais de vídeo provenientes da captura de cenas reais, além de proporcionar representação de vídeo conveniente à diversos tipos de transmissões. Com seu desempenho com taxas de bits até 50% menor que o antecessor MPEG-2 e com a mesma qualidade, sua difusão no meio digital foi gigantesca, sendo adotado como padrão principal na indústria de telecomunicações.

“O H.264 possui uma série de vantagens em relação a seus antecessores, com respeito às etapas de predição, transformação e codificação”. (MANUEL, 2007, pág. 4). Estas etapas serão analisadas nas próximas páginas.

A Figura 11 apresenta o diagrama de blocos simplificado do codificador e decodificador utilizados no padrão H.264.





Fonte: AGOSTINI, 2007.

Nota-se que este diagrama é bem parecido com o apresentado na Figura 9. Porém, no padrão o H.264 o módulo de Predição interquadros, na Figura 11, é composto pela estimação de movimento (ME – *Motion Estimation*) e pela compensação de movimento (MC – *Motion Compensation*), e é aqui onde se encontra a maior parte das inovações e dos ganhos obtidos neste padrão sobre os demais anteriores a ele, agregando operações de grande complexidade computacional a este módulo. (AGOSTINI, 2007). A estimação e compensação de movimento serão abordados separadamente e com maior detalhamento nesta seção.

O módulo de predição intraquadros, os blocos de transformadas diretas e inversas (T e TI), quantização direta e inversa (Q e QI) e o módulo da Codificação por Entropia tem a mesma função descrita no Codec da Figura 9, o mesmo ocorre para o decodificador.

Pode-se destacar também algumas vantagens do padrão H.264 com relação à predição intraquadros, codificação por transformada e codificação por entropia.

A predição intraquadros na estrutura do H.264 é definida por técnicas em que as amostras do sinal de vídeo são agrupadas em macroblocos de 16x16 pixels. A predição existe como uma unidade básica de identificar informações de macroblocos já identificados e pertencentes a um mesmo quadro. (MANUEL, 2007).

A codificação por transformada visa possibilitar a codificação do sinal residual, onde a redundância espacial existente no erro de predição é reduzida. Diferente dos padrões MPEG-1 e MPEG-2, o H.264/AVC não aplica a DCT em blocos com 8x8 pixels, mas sim transformadas inteiras reversíveis de tamanhos variáveis de 8x8, 4x4 a 2x2 amostras. (MANUEL, 2007). A partir disto, permite-se ao compressor codificar melhor o resíduo de predição quando são selecionados blocos de tamanho menores que 8x8,

garantindo uma melhor compactação do sinal devido à menor variação nas regiões pequenas de um quadro.

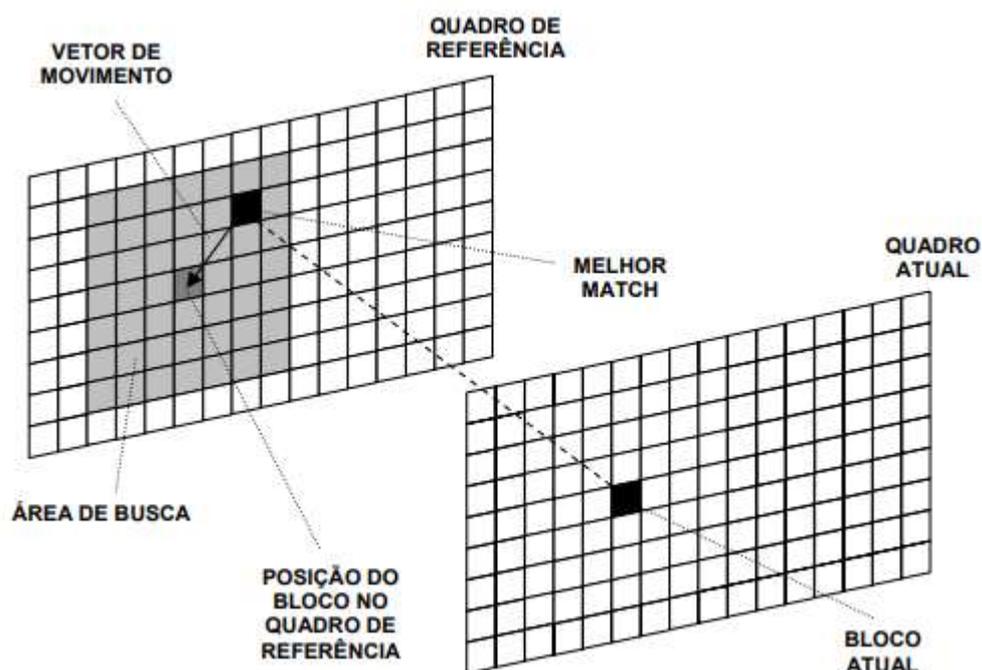
Na codificação por entropia o H.264/AVC possui um alto grau de adaptação às fontes, aumentando bastante a eficiência de codificação deste módulo, incorporando um modelamento de contexto do sistema baseado em dois tipos de codificação: de comprimento variável e aritmética binária adaptativa. Esse sistema analisa a hipótese de que a estatística dos sinais seria estacionária e com códigos de comprimento variável de tamanho para os elementos sintáticos normatizados.

### **4.3. Módulo de Estimação de Movimento (ME)**

A Estimação de Movimento (ME), é presente apenas nos codificadores do padrão H.264/AVC e juntamente com o Compensador de Movimento (MC), formam a predição interquadros, que é responsável pela maior complexidade computacional de um codificador H.264/AVC. (PURI, 2004). Este elevado custo operacional é devido a diversas novas ferramentas que foram inseridas nesta etapa, cujo objetivo foi aumentar a potência de compressão, em relação aos padrões anteriores de compressão de vídeo. (WIEGAND, 2003).

A estimação de movimento de um bloco busca encontrar, no quadro de referência, uma amostra que tem maior proximidade com o bloco do caso atual, cujo objetivo é diminuir a redundância na transmissão de quadros. A Figura 12 pode descrever este processo.

Figura 12 - Processo de estimação de movimento.



Fonte: PORTO, 2008.

“Em cada bloco do quadro atual é gerado um vetor de movimento que corresponde à posição para qual o bloco obteve a melhor relação de similaridade no quadro de referência” (PORTO, 2008), ou seja, este processo tem por objetivo localizar, no quadro de referência, qual o bloco mais semelhante ao bloco do quadro atual. Esta busca é feita em uma área ao redor do bloco em análise, chamada de área de busca. Em seguida, algoritmos de similaridade são utilizados para fazer a escolha do quadro mais similar, assim que o novo bloco é encontrado, o vetor de movimento é gerado.

Com este processo, minimiza-se a energia do quadro residual obtido, aumentando o desempenho de compressão. Contudo, a busca pela maior similaridade entre os blocos é uma tarefa computacional intensiva.

A Figura 13 ilustra o processo que foi descrito acima, onde compara-se a subtração simples de dois quadros sequenciais de um vídeo, um com estimação de movimento e o outro sem. O resultado da subtração simples superior, nos mostra um quadro residual sem estimação de movimento. Já na subtração inferior, agora com estimação de movimento, obtêm-se um novo quadro residual. Pode-se perceber que, na primeira operação, uma maior área escura ou clara, onde se tem diferenças negativas e positivas entre os quadros, respectivamente, e a área cinzenta, onde a diferença é nula. Quanto maior as áreas escuras ou claras, maior a quantidade de energia existente no quadro residual. Portanto, é visível

na segunda operação, com estimação de movimento, que as áreas escuras ou claras são menores, deixando uma menor energia residual no quadro resultante, cumprindo seu propósito.

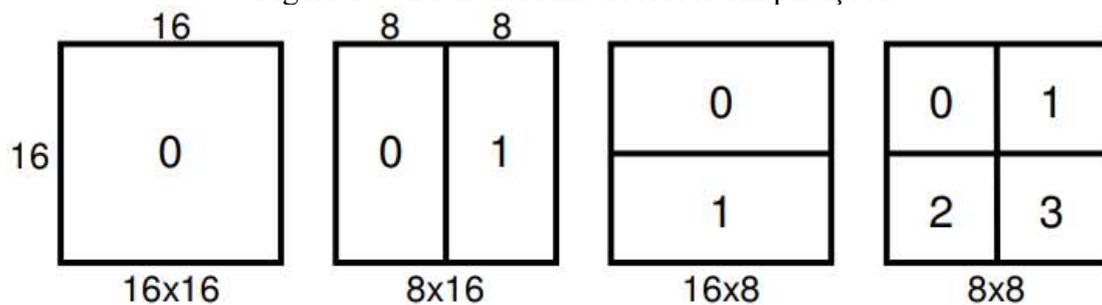
Figura 13 - Comparação entre quadro com e sem Estimação de Movimento.



Fonte: SCURI, 2002.

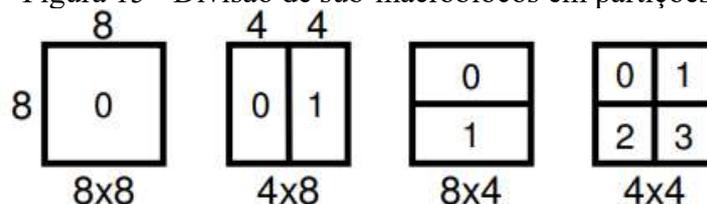
Como dito anteriormente, a principal inovação do padrão H.264/AVC com relação a ME e MC é a utilização de blocos de tamanhos variáveis. Então, o padrão H.264/AVC usa partições de macroblocos e partições de sub-macro blocos ao invés de gerar um vetor de movimento para cada macro bloco. Logo, cada uma dessas partições pode ter seu vetor de movimento, as partições de macroblocos tem valores de 16x16, 8x16, 16x8 e 8x8, apresentado na Figura 14. Caso a divisão escolhida seja 8x8, cada partição, os sub-macro blocos, podem ser divididas novamente, com valores de 8x8, 8x4,4x8 e 4x4, de acordo com a Figura 15. (RICHARDSON, 2003)

Figura 14 - Divisão de macroblocos em partições.



Fonte: AGOSTINI, 2007.

Figura 15 - Divisão de sub-macroblocos em partições.



Fonte: AGOSTINI, 2007.

A estimação de movimento é um extenso processo do padrão H.264/AVC, com diversos parâmetros que não serão aqui abordados, mas que podem ser analisados mais a fundo em Agostini (2007).

Existem diversas alternativas de algoritmos de busca de estimação de movimento, que podem ser classificados por uma busca completa, na qual realiza uma exaustiva comparação do bloco atual com todos os blocos candidatos dentro da janela de busca, encontrando um ótimo resultado para cada um dos blocos, em contrapartida a um alto custo computacional. Exemplos desses algoritmos de busca completa são: linha a linha e em espiral, e podem ser encontrados em Porto (2008).

Têm-se também a busca rápida, onde o número de operações para a estimação de movimento é reduzido, onde somente alguns blocos candidatos são comparados. Alguns exemplos desses algoritmos são: *Orthogonal Search*, *Four Step Search*, *Binary Search*, dentre outros.

#### 4.4. Compensação de Movimento (MC)

O processo de Compensação de Movimento opera de forma conjunta com a estimação de movimento que foi apresentada anteriormente. Na etapa de ME, é localizado o melhor casamento entre os macroblocos dentre os quadros de referência em uma região próxima, gerando um vetor de movimento. A partir do vetor de movimento gerado, encontrar os blocos que foram casados, na memória dos quadros que foram codificados anteriormente e montar o quadro de predição. No codificador, este quadro predito será subtraído do quadro atual, gerando o quadro de resíduos que passará pelos blocos de transformada. Já na decodificação os resíduos gerados pela transformada inversa são somados ao resultado da compensação. (AZEVEDO, 2006).

Diferente da Estimação de Movimento, a compensação de movimento está presente tanto no codificador, quanto no decodificador. Sendo que no codificador a MC atua de forma complementar à ME, podendo ser simplificada à medida que o processo ME

também é simplificado. No decodificador, a compensação de movimento deve ser completa de forma a reorganizar os macroblocos.

Como a compensação de movimento atua conforme o processo de estimação é executado, a MC deve atender os seguintes parâmetros, de acordo com Agostini (2007) e Azevedo (2006).

- Tratar os múltiplos tamanhos das partições dos macroblocos;
- utilizar múltiplos quadros de referência, tanto quadros passados como futuros;
- tratar os vetores de movimento que apontam para fora do quadro;
- interpretar corretamente os vetores de movimento criados na ME;
- no decodificador, reconstruir os macroblocos com precisão de 1/4 de pixels para os vetores de movimento de luminância;

A compensação de movimento não será abordada a fundo neste texto, podendo ser encontrada com mais detalhes em Azevedo (2006).

## 5. Padrão H.266/VVC

O grupo chamado *Joint Video Experts Team* (JVET), desenvolveram entre os anos de 2015 e 2020 o novo padrão de compressão de vídeo, *Versatile Video Coding* (H.266/VVC) (CHEN, 2020). A ideia principal é oferecer uma eficiência de compressão maior que o seu antecessor, o H.265/HEVC, com um *bitrate* de 50% menor, com a mesma qualidade de saída. Dito isto, este capítulo abordará o mais novo padrão H.266/VVC, explicitando suas otimizações em relação aos padrões anteriores e às novas ferramentas impostas a ele.

Quando seu modelo de testes foi lançado em 2020 e comparado com os demais padrões existentes, o VVC conseguiu superar seus antecessores (TAKAMURA, 2019). Porém, tudo no meio digital tem um preço, em troca de uma melhor eficiência de compressão, o H.266 requer um aumento significativo no quesito trabalho computacional.

Assim, neste novo padrão, foram criadas diversas novas ferramentas e otimizadas outras já existentes no HEVC. Pode-se destacar, de forma sucinta, as principais delas e suas características no padrão VVC abaixo:

- **Predição Intraquadro** – No padrão HEVC, obteve-se 35 modos de predição intra, já no novo padrão VVC têm-se 67 modos diferentes, permitindo um maior ângulo para a refinação da predição. O VVC também realiza três novas técnicas de se executar a predição intraquadro em um bloco: a *Cross-Component Linear Mode*, onde usando um modelo linear, as amostras de luma reconstruídas são usadas para realizar a predição das amostras de *Chroma* para a mesma CU. O *Intra Sub-Partitions*, onde todas as amostras de Luma podem ser divididas em duas ou quatro subpartições, compartilhando o mesmo modo intra, no entanto, o processamento é feito gradualmente, cada uma utiliza as amostras reconstituídas nos quadros anteriores para gerar a predição da subpartição do quadro atual. A *Matrix-based Intra Prediction*, seleciona uma linha de amostras próximas reconstituídas como vetores de entrada. Por fim, o VVC aumenta as amostras que são usadas como referência, possibilitando o uso de diversas linhas de referências, melhorando a qualidade da predição.
- **Predição Interquadros** – Resumidamente, aqui a técnica chamada de *Generalized BI-Prediction* é utilizada, onde o modo de predição bilateral é aumentado para suportar até cinco pesos. O processo de estimação de

movimento também foi otimizado, podendo ser fracionado a uma predição de 1/16 nas amostras de Luma. Contudo, têm-se também a nova ferramenta de predição de movimento *Affine*, capaz de trabalhar com os movimentos irregulares presentes nos vídeos.

- **Transformada** – Nesta etapa o VVC pode utilizar blocos de maior resolução, explorando a redundância espacial, em seguida utiliza-se uma nova transformada secundária de menor resolução.
- **Codificação por Entropia** – Neste bloco, é aplicado uma otimização do *Context-Adaptative Binary Arithmetic Coding* (CABAC), esta técnica pode ser encontrada em Sequeira (2019).
- **Filtros em Loop** – Oferendo uma qualidade subjetiva melhor, nesta etapa têm-se a implementação de três filtros em Loop: *Deblocking Filter*, *Sample Adaptive Offset* (SAO) e *Adaptive Loop Filter* (ALF). Onde o último está presente apenas no padrão VVC, no qual utiliza um trabalho de classificação de blocos para selecionar entre os vinte e cinco filtros diferentes. Este filtro é aplicado somente em blocos de resolução mínima.

A codificação do Padrão H.266 foi desenvolvida em três configurações, para alcançar os requisitos impostos pela ITU. A primeira delas é a configuração *All Intra*, onde todos os quadros são codificados utilizando apenas a predição intraquadro, logo, não existem dependências entre quadros codificados.

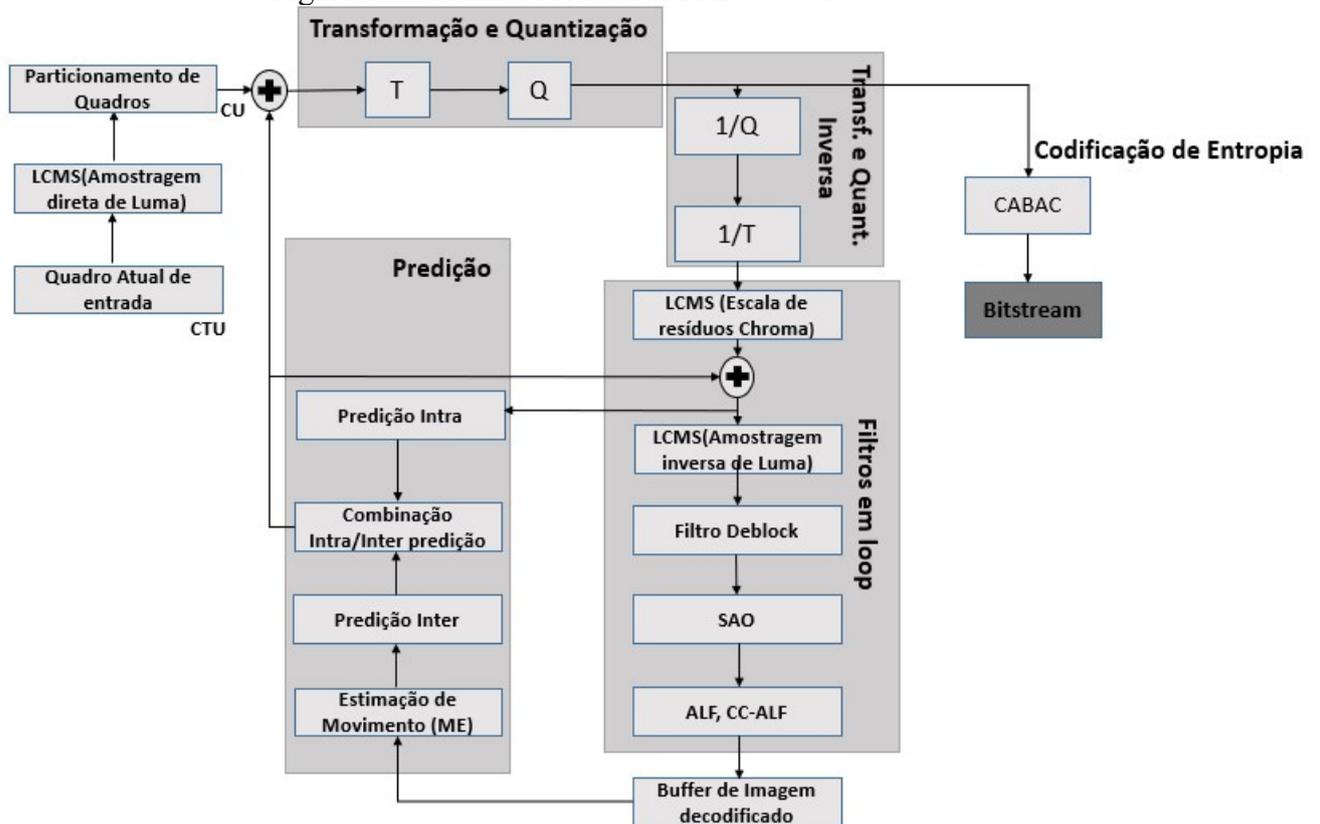
Em seguida têm-se a configuração *Low-Delay*, em que apenas o primeiro quadro do vídeo é codificado através da predição intraquadros, os quadros seguintes são codificados com a predição interquadros, criando uma dependência em relação aos quadros anteriores. Essa configuração apresenta uma boa taxa de compressão, visto que os quadros são codificados na mesma ordem em que são exibidos para o usuário.

Por fim, há a configuração *Random Access*, o codificador marca alguns pontos de salvamento durante a codificação de um quadro, utilizando um contador de intervalo (*Intra Period*), este por sua vez determina quantos e quais quadros serão codificados com a predição interquadros até que um quadro seja codificado somente com a predição intraquadro. Através desta técnica, minimiza-se a dependência criada para alguns quadros, tornando o sistema de transmissão mais sólido.

A Figura 16 ilustra o diagrama de blocos do codificador do novo padrão VVC. Este codificador é composto por seis blocos principais, são eles: Particionamento de Imagem;

Predição; Transformação e Quantização Direta e Inversa; Filtros em Loop e Codificação por Entropia. A seguir são destacadas as principais mudanças e otimizações deste novo padrão.

Figura 16 - Codificador Padrão H.266/VVC.



Fonte: Adaptado de HAMIDOUICHE, 2021.

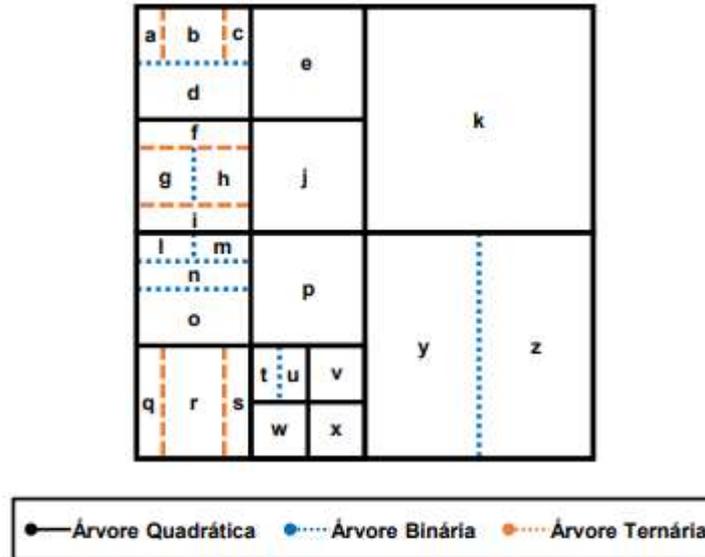
## 5.1. Particionamento de Imagem

No padrão anterior, HEVC, uma das grandes inovações foi na estrutura de particionamento flexível. O padrão VVC estende ainda mais essa estrutura, adicionando mais possibilidades de particionamentos, permitindo uma representação melhor dos objetos em cena.

No padrão VVC, cada quadro de entrada pode ser particionado em blocos de tamanho máximo 128x128 amostras. Onde, cada bloco é chamado de *Coding Tree Unit* (CTU), e estes mesmos blocos podem ainda ser particionados em blocos menores, denominados *Coding Units* (CUs), podendo ter tamanhos entre 64x64 até 4x4 amostras. Esta nova estrutura de particionamento é chamada de *Quadtree with nested Multi-Type* (QTMT).

A Figura 17 mostra a estrutura QTMT com uma CTU de tamanho 128x128 amostras divididas em várias CUs de diferentes níveis de QT e MTT.

Figura 17 - Particionamento QTMT para CTU de tamanho 128 x128.



Fonte: GONÇALVES, 2021.

No particionamento QTMT, a CTU atual é particionada em quatro CUs quadradas de tamanho igual, de acordo com a estrutura de particionamento de árvore quadrática (QT). Em seguida, cada uma das CUs geradas pode ser particionada novamente em outras CUs quadráticas usando QT, como em CUs retangulares através da técnica de estrutura *Multi-Type Tree* (MTT).

A estrutura de particionamento MTT pode ser usada para que cada CU seja dividida em até quatro tipos diferentes de árvores: Horizontal Binária (do inglês, *Binary Tree Horizontal* - BTH), Vertical Binária (do inglês, *Binary Tree Vertical* - BTV), Horizontal ternária (do inglês, *Ternary Tree Horizontal* - TTH) e Vertical ternária (do inglês, *Ternary Tree Vertical* - TTV). A divisão de árvore binária divide a CU atual em duas formas simétricas, já a divisão de árvore ternária divide a CU atual em outras três, cada nova CU com 25%, 50% e 25% do tamanho da CU atual, respectivamente. Este processo de particionamento pode ser aplicado tanto na vertical como na horizontal de um bloco. (CHEN, 2020).

A Figura 17 exhibe o particionamento QTMT para uma CTU com tamanho 128x128, dividida em diversas CUs utilizando estruturas QT e MTT. De acordo com a legenda, a

linha preta apresenta os particionamentos em árvore quadrática, já as linhas pontilhadas representam os particionamentos em árvores ternárias.

Através deste tipo de estrutura de particionamento QTMT, o padrão VVC possui uma maior capacidade dos tipos de particionamentos de blocos, adaptando esta ferramenta para os diferentes tipos de quadros e suas singularidades, oferecendo uma melhor eficiência de compressão.

## 5.2. Ferramentas de Codificação Intra

O princípio da codificação Intra aproveita a correlação espacial existente na textura da imagem. Para descorrelacioná-los, o padrão VVC utiliza uma série de ferramentas de codificação que são vinculadas ao particionamento e a um conjunto de modos de predição intra (do inglês, *Intra prediction modes* - IPMs). No particionamento de blocos, além das técnicas que foram abordadas na seção anterior, uma nova ferramenta denominada árvore dupla é introduzida, permitindo particionar árvores separadas entre luma e chroma, oferecendo um maior nível de liberdade para decisões de codificação dos blocos chroma.

As referências de predição intra são aprimoradas no padrão VVC para maior eficiência de controle das amostras vizinhas com ruído e menos correlacionadas. Uma ferramenta de seleção de referência, chamada de Referência de múltiplas Linhas (do inglês, *Multiple Reference Lines* – MRL) foi introduzida para permitir ao codificador escolher entre três linhas de referência e sinalizar a melhor delas com base nas características do bloco. (HAMIDOUICHE, 2021).

Além desta ferramenta, têm-se: *Cross-Component Linear Mode*, *Intra Sub-Partitions*, *Matrix-based Intra Prediction*. Que foram explicadas de forma sucinta no início deste capítulo para esta predição.

## 5.3. Ferramentas de Codificação Inter

Como citado anteriormente, a predição interquadros tem por finalidade identificar e tratar as redundâncias temporais encontradas entre os quadros. Esta etapa é bastante similar ao padrão anterior, HEVC, utilizando o uso de quadros de referências já codificados para realizar os algoritmos de buscas.

Os quadros de referência são então armazenados em dois *buffers*, denominados de lista L0 e L1. A lista L0 armazena os quadros de referência temporalmente passados, e a

lista L1 armazena os quadros de referência temporalmente futuros. Cada uma dessas listas pode armazenar até 16 quadros de referência. (GONÇALVES, 2021).

Seguindo o mesmo princípio dos padrões de codificadores anteriores, a predição interquadros do padrão VVC possui duas principais etapas, a Estimação de Movimento (ME) e a Compensação de Movimento (MC). Ambas já tiveram suas características e definições exibidos nas seções anteriores, e no padrão VVC funcionam de maneira muito similar, mas vale destacar algumas otimizações que o novo padrão trouxe na etapa de ME.

Primeiramente, para determinar a similaridade entre os blocos, o VVC analisa um parâmetro chamado de taxa de distorção (do inglês, *Rate-Distortion* – RD), buscando nos quadros de referência por blocos que possam reduzir o custo de RD. Esta técnica, busca otimizar a quantidade de distorção em relação à quantidade de dados necessários para que o vídeo possa ser codificado. (HOANG; LONG; VITTER, 1998).

Neste caminho, a estimação de movimento no VVC pode ser definida em três etapas centrais: a primeira delas é a predição unilateral, onde cada quadro de referência é avaliado para definir a área de busca para encontrar os blocos com maior similaridade. Após terminada a avaliação dos quadros de referência, o bloco que obtiver o menor custo RD é selecionado das listas L1 e L0.

A segunda etapa é a predição bilateral, como já citada anteriormente, esta predição analisa os melhores vetores de movimento na etapa anterior, melhorando o bloco candidato. Os melhores blocos das listas L1 e L0 são somados, gerando um bloco com custo de RD menor que os dois candidatos escolhidos na etapa unilateral. No VVC, a precisão dos vetores de movimento é de 1/16 amostras para os componentes luma e 1/32 amostras para os componentes de chroma.

A terceira etapa é a predição *Affine*, esta ferramenta é uma novidade em relação aos padrões anteriores. Resumidamente, esta predição visa mapear alguns movimentos dos objetos nos quadros, tais como rotação, cisalhamento e *zooming*. Quando uma CU é codificada através da predição *Affine*, é dividida em sub-blocos 4x4, cujo os vetores de movimento são derivados dos parâmetros *Affine* da CU, estes parâmetros são definidos através dos vetores de movimento do ponto de controle, estrategicamente posicionados nas bordas superior esquerda e superior direita. Neste texto, não será abordado esta predição a fundo, mas pode ser encontrada com mais detalhes em Gonçalves (2021).

#### 5.4. Filtros em *Loop*

O particionamento de imagens e as etapas de quantização utilizadas no padrão VVC podem causar eventos indesejados de codificação, como descontinuidade entre os blocos, ruídos, textura e suavização das bordas. Para mitigar estes eventos, quatro filtros em loop são implementados em VVC, aumentando a eficiência geral da codificação. Sendo eles: Filtro de desbloqueio (do inglês, *Deblocking Filter* – DBF), deslocamento de amostra adaptativo – SAO, filtro de loop adaptativo (do inglês, *Adaptative Loop Filter* –ALF) e filtro de loop adaptativo de componente cruzado (do inglês, *Cross-Component Adaptive loop Filtering* – CC/ALF). (Karczewicz, 2021).

Além disso, têm-se uma nova ferramenta chamada de Mapeamento Luma com escalamento Chroma (do inglês, *Luma Mapping with Chroma Scaling* – LMCS), que executa o mapeamento das componentes de luma para o sinal de predição interquadros e escala de chroma para os resíduos gerados após as etapas de transformação e quantização inversas.

O filtro DBF é aplicado nos limites dos blocos para reduzir os eventos de bloqueio de informação. O filtro SAO é então aplicado nas amostras desbloqueadas, classificando as amostras desbloqueadas em diferentes categorias e, para cada uma delas, um valor de deslocamento recuperado pelo decodificador de entropia é adicionado nas amostras.

O filtro ALF realiza a filtragem adaptativa com o objetivo de minimizar o erro quadrático médio (MSE) entre as amostras originais e as reconstituídas. Por fim, o filtro CC/ALF usa amostras de luma para gerar uma correção nas amostras de chroma, sendo realizado paralelamente ao filtro ALF. (HAMIDOUCHE, 2021)

#### 5.5 Resultados Comparativos

Neste tópico serão apresentados dados experimentais resultados da utilização dos padrões de codificação de vídeos H.264/AVC e do padrão H.266/VVC. Para efeitos de comparação, serão consideradas as sequências de vídeos propostas no documento *Common Test Condition* (CTC) de ambos os padrões, disponíveis em (TOURAPIS, 2009) (CHEN, 2020).

Abaixo, a Tabela 1 apresenta sequências de vídeos que podem ser utilizados para realização dos experimentos seguindo as recomendações do CTC, os vídeos são classificados de A à F de acordo com sua resolução. Percebe-se que a classe B utiliza 10

ou 8 bits para representação dependendo de sua classificação dentro de perfis e níveis definidos pelo padrão, porém, sempre com uma resolução FHD de 1920x1080 pixels. Logo, os vídeos que pertencem a esta classe podem ser utilizados por ambos os padrões H.264/AVC e H.266/VVC para o processo comparativo entre eles.

*Tabela 1 - Sequência de testes de acordo com o CTC de JVET VVC.*

<b>Classe</b>	<b>A (10-bit)</b>	<b>B (10-bit/8-bit)</b>	<b>C (8-bit)</b>	<b>D (8-bit)</b>	<b>F (8-bit)</b>
<b>Resolução</b>	3840x2160 (4k)	1920x1080 (FHD)	832x480	416x240	Variável
	Tango2(60fps)	MarketPlace (60fps, 10-bit)	RaceHorses (30fps)	RaceHorses (30fps)	ArenaofValor (60fps, 1920x1080)
	FoodMarket (60fps)	RitualDance (60fps, 10-bit)	BQMall (60fps)	BQSquare (60fps)	BasketballDrillText (50fps, 832x480)
<b>Sequência</b>	Campfire (30fps)	Cactus (50fps, 8-bit)	PartyScene (50fps)	BlowingBubbles (50fps)	SlideEditing (30fps, 1280x720)
	CatRobot (60fps)	BasketballDrive (50fps, 8-bit)	BasketballDrill (50fps)	BasketballPass (50fps)	SlideShow (20fps, 1280x720)
	ParkRunning3 (50fps)	BQTerrace (60fps, 8-bit)	-	-	-

Fonte: Adaptado de GROIS, 2020.

Para efeitos comparativos, uma das principais métricas para análise da qualidade de um vídeo digital utilizada é a relação sinal-ruído, ou PSNR (do inglês, *Peak Signal to Noise Ratio*). A PSNR pode ser definida como uma métrica que avalia a qualidade do vídeo, após o processo de codificação e decodificação, comparando o vídeo original com o vídeo processado adicionado ao ruído que é acrescentado no canal. Este parâmetro é medido em decibéis (dB) e pode ser calculado de acordo com a Equação 5.1.

$$PSNR = 20 \log_{10} \left( \frac{MAX}{\sqrt{MSE}} \right) \quad (5.1)$$

Onde, MAX é o valor máximo que um pixel pode representar. Considerando uma representação utilizando 8 bits por pixel, têm-se MAX = 255. O MSE corresponde à média do quadrado da diferença entre cada uma das amostras da imagem original e

amostra após o processo de compressão, ou seja, quanto maior o MSE pior a qualidade do vídeo após a compressão. É calculado através da Equação 5.2.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (R_{ij} - O_{ij})^2 \quad (5.2)$$

Em que, R e O são as matrizes que representam as amostras de pixels do quadro original e o quadro reconstituído, respectivamente. No que diz respeito à  $m$  e  $n$ , entende-se que representam o número de amostras que as matrizes R e O possuem. (GHANBARI, 2003)

Para comparação dos padrões, foram adotados os parâmetros recomendados sem variação de QP (28). Foram utilizados os *Softwares* de Referência para o padrão H.264/AVC (JM) e o VVC Test Model (VTM) para o padrão H.266/VVC, disponíveis em Kim (et al., 2013) e Chen (2020), respectivamente. Os experimentos foram executados sobre uma plataforma computacional de teste com processador Intel Xeon E5 2643 v1 com 16Gb de memória RAM ECC. O sistema utilizou o compilador GCC padrão através do Sistema Operacional Linux Mint 20.1.

Considerando o exposto, foram realizadas a partir dos vídeos fonte da classe B, exibidos na Tabela 1, simulações comparando a métrica de qualidade visual PSNR pelo tempo de codificação para cada um dos padrões H.264/AVC e H.266/VVC. Estes valores estão destacados na Tabela 2.

*Tabela 2 - Resultado comparativo entre os padrões AVC e VVC para PSNR pelo tempo de codificação.*

<b>Sequência</b>	<b>Métrica/Padrão</b>	<b>H.264/AVC</b>	<b>H.266/VVC</b>
MarketPlace	PSNR (dB)	34,32	38,9
	Tempo de Cod. (s)	2723	68841
RitualDance	PSNR (dB)	33,5	38,93
	Tempo de Cod. (s)	3115	81108
Cactus	PSNR (dB)	32,8	37,78
	Tempo de Cod. (s)	2850	70336
BasketballDrive	PSNR (dB)	33,79	39,11
	Tempo de Cod. (s)	3243	80653
BQTerrance	PSNR (dB)	34,01	38,07
	Tempo de Cod. (s)	2919	71555

Fonte: Autoria própria.

A partir da Tabela 2, pode-se verificar que para a sequência de vídeos utilizada nos testes, de acordo com a PSNR obtida, a qualidade visual do vídeo após a codificação foi maior no padrão H.266/VVC. Desta maneira, verifica-se que o novo padrão VVC possui em seus resultados menor inserção de perdas ou ruídos, obtendo um vídeo codificado mais fiel ao vídeo original. Porém, ao analisar o tempo de codificação em ambos os padrões, nota-se que o padrão H.264/AVC obteve um tempo bem menor de codificação em todos os vídeos analisados, isso se deve ao fato do codificador AVC ser bem menos complexo computacionalmente em relação ao padrão VVC, que por sua vez, demanda um alto desempenho computacional.

A Tabela 3 apresenta também uma comparação entre os dois padrões, exibindo a saída de dados em taxa de bits (*Bitrate*) pela complexidade computacional em termos do tempo de codificação em segundos.

*Tabela 3 - Resultado comparativo entre a taxa de Bitrate pelo tempo de codificação dos padrões AVC e VVC.*

Sequência	Métrica/Padrão	H.264/AVC	H.266/VVC
MarketPlace	<i>Bitrate</i> (kbps)	1960,1	366,2
	Tempo de Cod. (s)	2723	68841
RitualDance	<i>Bitrate</i> (kbps)	2399,3	370,6
	Tempo de Cod. (s)	3115	81108
Cactus	<i>Bitrate</i> (kbps)	2208,4	366,6
	Tempo de Cod. (s)	2850	70336
BasketballDrive	<i>Bitrate</i> (kbps)	1998,2	362
	Tempo de Cod. (s)	3243	80653
BQTerrance	<i>Bitrate</i> (kbps)	2103,5	369,3
	Tempo de Cod. (s)	2919	71555

Fonte: Autoria própria.

A partir da Tabela 3, pode-se entender que para a mesma sequência de vídeos, o padrão H.266/VVC obteve um *Bitrate* muito inferior em relação ao padrão H.264/AVC. Isso conclui que o poder de compressão do novo padrão VVC é muito superior ao padrão AVC, porém com um tempo de codificação inversamente proporcional, demandando uma grande capacidade computacional da máquina utilizada. Como exemplo, referindo-se ao vídeo fonte MarketPlace, o mesmo detém um tempo total de codificação de 45 minutos

para o padrão AVC e 19 horas para o padrão VVC, ou seja, um aumento de aproximadamente 20 vezes no tempo total de codificação para o mesmo vídeo.

Por fim, vale destacar que estes testes foram realizados com os vídeos de classe B, cujo resolução máxima que pode ser obtida pela sequência de vídeos é o FHD. Porém, a classe A de vídeos seria mais adequada para este comparativo, devido a sua resolução em 4k. Com esta resolução em 4k pode-se facilmente deduzir que o tempo de codificação e o *Bitrate* de saída em ambos os padrões seria superior ao apresentado nas Tabelas 3.

Por exemplo, para o vídeo fonte MarketPlace apresentado na Tabela 3, é possível estimar que os valores seriam aproximadamente de acordo com a Tabela 4. Logo, conforme os resultados obtidos, é possível afirmar que quanto maior a resolução do vídeo maior o tempo de codificação e maior o *bitrate* gerado para ambos os padrões, chegando a elevar o tempo de codificação do mesmo vídeo fonte para 3 e 76 horas aproximadamente, para os padrões AVC e VVC, respectivamente.

*Tabela 4 - Resultado comparativo entre os padrões AVC e VVC utilizando resolução 4K*

Sequência	Métrica/Padrão	H.264/AVC	H.266/VVC
MarketPlace	<i>Bitrate</i> (kbps)	7840,4	1464,8
	Tempo de Cod. (s)	10892	275364

Fonte: Autoria própria.

## 6. Considerações Finais

Este trabalho apresentou um estudo comparativo entre as estruturas que compõem os padrões H.264/AVC e H.266/VVC. O foco desta comparação se dá em apresentar as características do padrão H.264/AVC que o tornou um divisor de águas em termos de tecnologia para codificadores de vídeos em seu lançamento. Por outro lado, abordando as demandas atuais de resoluções cada vez melhores, têm-se o mais novo padrão H.266/VVC que trouxe um novo perfil para o momento atual.

O padrão H.264/AVC trouxe em seu codificador, no módulo de predição, as etapas de estimação e compensação de movimento, as etapas de ME e MC potencializaram em muito a capacidade de compressão de vídeo, gerando uma menor taxa de bits para representar uma maior informação.

Por outro lado, para atender às demandas atuais, obtendo uma eficiência de compressão superior aos seus antecessores possibilitando serviços com resoluções de até 16k, o padrão H.266/VVC trouxe diversas melhorias em seu Codec, como o particionamento de blocos, trazendo uma nova estrutura de particionamento de blocos QTMT. Têm-se também a implementação do módulo de filtros capaz de reduzir ainda mais os efeitos de ruído e textura inerentes ao processo. As etapas de predições foram otimizadas, com destaque para a nova ferramenta de predição *Affine*, que busca pontuar as transformações não translacionais presentes nos vídeos digitais. Entretanto, o padrão H.266/VVC apresentou a necessidade de alta capacidade computacional em sua codificação, sobretudo no módulo ME, que no VVC chega a representar quase 60% do tempo total de codificação.

Este estudo trouxe pontos importantes dos padrões H.264/AVC e H.266/VVC, abordando novos módulos e otimizações do novo padrão. Para trabalhos futuros, sugere-se o estudo de algoritmos que visam reduzir o esforço computacional, sobretudo no módulo ME do padrão H.266/VVC.

Por fim, o autor acredita que o padrão VVC trará uma nova realidade, porém o padrão AVC ainda perdurará por um longo período, devido a sua compatibilidade com diversos sistemas e difusão no mercado atual.

## REFERÊNCIAS

- AGOSTINI, Luciano. **Desenvolvimento de Arquiteturas de Alto Desempenho Dedicadas à Compressão de Vídeo Segundo o Padrão H.264/AVC**. Tese de Doutorado em Ciência da Computação – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.
- ZAGHETTO, Alexandre. **Compressão de documentos compostos utilizando o H.264/AVC-INTRA**. XXVII, 177 f. Tese (Doutorado em Engenharia Elétrica) - Universidade de Brasília, Brasília, 2009.
- WYSZECKI, G.; STILES, W. S. **Color Science: Concepts and methods, quantitative data and formulae**. 2<sup>nd</sup>. Ed.: Wiley, 2000.
- GONZALEZ, R.; WOODS, R. **Processamento de Imagens digitais**. São Paulo: E, Blucher, 2003.
- RICHARDSON, I. **H.264/AVC and MPEG-4 Video Compression – Video coding for Next – Generation Multimedia**. Chichester: Jhon Wiley and Sons, 2003.
- ALBUQUERQUE, M. P.; ALBUQUERQUE, M. P. **Processamento de imagens: Métodos e Análises**. Rio de Janeiro. Centro brasileiro de pesquisas físicas – CBPF/MCT, 2009.
- BETETTO, Luciana Aparecida de Oliveira. **Método para compressão de imagens digitais fundamentado em procedimentos de Huffman e Wavelets**. 2005. 182 f. Dissertação (Mestrado em Ciências Exatas e da Terra) - Universidade Federal de São Carlos, São Carlos, 2005.
- SCURI, A. E. **Fundamentos da imagem digital**. Rio de Janeiro, 2002. Disponível em < <https://web.tecgraf.puc-rio.br/~scuri/download/fid.pdf> > Acesso em: 28 out. 2021.
- SEQUEIRA, Luis Frederico Tokumoto. **Exploração de técnicas de redução de potência em arquitetura de hardware da transformada discreta do cosseno 2-D para o padrão H.265/HEVC**. 73 f. Dissertação (Mestrado em Engenharia Eletrônica e Computação) - Universidade Católica de Pelotas, Pelotas, 2019.
- SAYOOD, K. **Introduction to Data Compression**. [S.l.]: Morgan Kuffmann Publishers, 2000.
- MONTEIRO, Estêvão Chaves; SANTOS, Lucas Alberto Souza; TECNOLÓGICA, Infraestrutura. **Modernização dos Sistemas Audiovisuais do Governo Baseada em Software Livre**. ConSerpro - Congresso Serpro de Tecnologia e Gestão aplicadas a Serviços Públicos. Regional São Paulo - 3-6 de dezembro de 2013.
- WIEGAND, T. et al. Overview of the H. 264/AVC video coding standard. **IEEE Transactions on circuits and systems for video technology**, [S.l.], v.13, n.7, p.560–576, 2003.

KUHN, P. **Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation**. Boston: Kluwer Academic Publishers, 1999.

GOEBEL, Jones William. **Solução Arquitetural para a Decodificação Intraquadro do Padrão de Codificação Av1**. 2019. 117 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2019.

WU, Z., SHA, J., WANG, Z., LI, L., and GAO, M. Gao. **An improved scaled DCT architecture**. IEEE Transactions on Consumer Electronics, vol. 55, no. 2, pp. 685–689, May 2009

ROSA, Vagner Santos. **Arquiteturas de Hardware Dedicadas para Codificadores de Vídeo H.264** – Filtragem de Efeitos de Bloco e Codificação Aritmética Adaptativa ao Contexto [manuscrito], 2010.

**Committed to Connecting the World**. ITU, 2021. Disponível em: <itu.int/em/ITU-T >- Acesso em: 08 de Ago 2021.

HUSEMANN, et al. **Recomendações técnicas da Equipe RFP-03 para o SBTVD: codificador e decodificador escalável MPEG-2**. São Leopoldo: Unisinos, 2006.

CHEN, Y, K.; PENG, W. H. **Implementation of rea-time MPEG-4 FGS encoder**. In: IEEE PACIFIC-RIM CONFERENCE ON MULTIMEDIA, 3. 2002, Hsinchu: PCM, 2002.

BENDER, Isis. **Redução de Complexidade na Compressão de Vídeo: Relevância e Aplicação no Codificador AV1**. Institucional UFPEL,2020. Disponível em: [https://ppgc.ufpel.edu.br/documentos/qualificacao/Isis\\_Duarte\\_Bender\\_Reducao\\_de\\_complexidade\\_na\\_Compressao\\_de\\_Video\\_Relevan\\_aRr958P.pdf](https://ppgc.ufpel.edu.br/documentos/qualificacao/Isis_Duarte_Bender_Reducao_de_complexidade_na_Compressao_de_Video_Relevan_aRr958P.pdf) Acesso em: 10 de Nov 2021.

MANUEL, Edson T. M. **Codificação de Vídeo H.264 – Estudo de Codificação Mista de Macroblocos**. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, 2008.

PURI, A. et al. **Video Coding Using the H.264/MPEG-4 AVC Compression Standard**. Elsevier Signal Processing: Image Communication, [S.l.], n.19, p.793-849, 2004.

PORTO, R. **Desenvolvimento Arquitetural para Estimação de Movimento de Blocos de Tamanhos Variáveis Segundo o Padrão H.264/AVC de Compressão de Vídeo**, 2008. 96f. Dissertação (Mestrado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre.

RICHARDSON, I. **H.264/AVC and MPEG-4 Video Compression – Video Coding for Next Generation Multimedia**. Chichester: John Wiley and Sons, 2003.

AZEVEDO, A. P. **MoCHA: Arquitetura Dedicada para a Compensação de Movimento em Decodificadores de Vídeo de Alta Definição, seguindo o Padrão**

**H.264.** Dissertação (Mestrado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre, RS, 2006.

GONÇALVES, Paulo Henrik Ribeiro. **Um Esquema Rápido Baseado em Aprendizado de Máquina para a Predição Interquadros do Codificador de Vídeo VVC.** 90 f. Dissertação (Mestrado em Ciência da Computação) – Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2021.

CHEN, J. **Algorithm description for Versatile Video Coding and Test Model 8 (VTM).** JVET-Q2002. ed. Bélgica: [s.n.], 2020.

HAMIDOUCHE, W. et al. **Versatile Video COding Standard: A Review from Coding Tools for Consumers Deployment.** IEEE Consumer Electronics Magazine, 2021.

HOANG, D. T.; LONG, P. M.; VITTER, J. S. **Efficient Cost Measures for Motion Estimation at Low Bitrates.** IEEE Transactions on Circuits and Systems for Video Technology, [S.l.], v.8, n.4, p.488–500, 1998.

KARCZEWICZ, M. et al. **“Vvc in-loop filters”** IEEE Transactions on Circuits and Systems for Video Technology, pp. 3907 – 3925, 2021.

SULLIVAN, G. et al. **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, 2012.

GHANBARI, M. 2003. **Standard Codecs: Image Compression to Advanced Video Coding.** [S.l.]: Institution Electrical Engineers, 2003.

GROIS, D. **Comparison Assessment of Emerging EVC and VVC Video Coding Standards with HEVC and AV1.** Comcast, 2020.

CHEN, J. et al. **Algorithm description for Versatile Video Coding and Test Model 11 (VTM 11).** JVET, 2020.

TOURAPIS, A. M. **H.264/14496-10 AVC Reference Software Manual.** Microsoft Corporation, 2009.

KIM, I.-K. et al. **HM: High Efficiency Video Coding (HEVC) Test Model Encoder Description.** JCT-VC-L1002. Geneva: s.n., 2013.