



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

IURY GLABSON OLIVEIRA SILVA

**AVALIAÇÃO EXPERIMENTAL DA COMPRESSÃO DE VÍDEOS DE
ESTRUTURAS CIVIS UTILIZANDO TÉCNICAS DE ANÁLISE MODAL**

**BELÉM
2022**



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

IURY GLABSON OLIVEIRA SILVA

**AVALIAÇÃO EXPERIMENTAL DA COMPRESSÃO DE VÍDEOS DE
ESTRUTURAS CIVIS UTILIZANDO TÉCNICAS DE ANÁLISE MODAL**

Trabalho de Conclusão de Curso apresentado
para obtenção do grau de Bacharel em Ciência
da Computação.

Orientador: Prof. Dr. Reginaldo Cordeiro Dos
Santos Filho

Coorientador: Dr. Moisés Felipe Mello da Silva

**BELÉM
2022**

SILVA, IURY

avaliação experimental da compressão de vídeos de estruturas civis utilizando técnicas de análise modal/ IURY GLABSON OLIVEIRA SILVA. – BELÉM, 2022.

48 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Reginaldo Cordeiro Dos Santos Filho

Coorientador: Dr. Moisés Felipe Mello da Silva

Monografia – UNIVERSIDADE FEDERAL DO PARÁ

INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS

CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO, 2022.

1. STRUCTURAL HEALTH MONITORING. 2. VIDEO COMPRESSION. 3. MODAL ANALYSIS. 4. VIDEO MOTION PROCESSING I. Título.

IURY GLABSON OLIVEIRA SILVA

**AVALIAÇÃO EXPERIMENTAL DA COMPRESSÃO DE VÍDEOS DE
ESTRUTURAS CIVIS UTILIZANDO TÉCNICAS DE ANÁLISE
MODAL**

Trabalho de Conclusão de Curso apresentado
para obtenção do grau de Bacharel em Ciência
da Computação.

Data da Defesa: 15 de Dezembro de 2022

Conceito: Excelente

Banca Examinadora

**Prof. Dr. Reginaldo Cordeiro Dos Santos
Filho**

Faculdade de Computação - UFPA
Orientador

**Prof. Dr. Claudomiro de Souza de Sales
Junior**

Faculdade de Computação - UFPA
Membro da Banca

Prof. Dr. Adam Dreyton Ferreira dos Santos

Faculdade de Sistemas de Informação -
UNIFESSPA
Membro da Banca

BELÉM
2022

*Este trabalho é dedicado aos meus pais, Gleidson e Olieonay,
e ao meu avô, Orlandino (in memoriam), que segue sendo minha inspiração.*

AGRADECIMENTOS

Agradeço primeiramente aos meus amados pais, por estarem sempre presentes em minha vida e por todo apoio prestado durante o período do curso.

Ao meu orientador, prof. Reginaldo, por todos os ensinamentos e pela assistência dada ao decorrer da elaboração deste documento e da minha jornada acadêmica. Agradeço também ao meu coorientador Moisés pela ajuda prestada no desenvolvimento deste trabalho.

A toda minha família e amigos, que me incentivaram e compreenderam a minha ausência enquanto eu me dedicava à universidade.

Aos amigos que fiz dentro da universidade, que sempre estiveram ao meu lado, pela amizade incondicional e pelo apoio demonstrado ao longo de todo o curso.

- Antonio Melgacino de S. Neto
- Aimeê Miranda Ribeiro
- Eduardo Gil S. Cardoso;
- Gabriel Afonso P. da Silva;
- Gabriela S. Maximino;
- George Felipe de M. Silva
- Giovanne César O. de Souza;
- Guilherme Eiji E. Hantani
- João Lucas D. Silva;
- Karen Sulliany S. da Silva
- Lucas Mesquita R. Ferreira;
- Marco Aurélio Lima do N. Junior
- Rafael da S. Lisboa;
- Igor Matheus S. Moreira

*“Pois as crianças são inocentes e amam
a justiça, enquanto a maioria de nós é corrompida
e naturalmente prefere o perdão.”
(G. K. Chesterton)*

RESUMO

Na área de monitoramento da saúde de estruturas, a análise modal é muito utilizada para verificar a condição física das estruturas, a partir da extração de características relevantes para identificação de dano. Para isso, geralmente as propriedades modais (frequências naturais, modos de vibração e amortecimentos modais) são utilizadas, pois as mudanças nas características físicas da estrutura também podem ser vistas nas propriedades modais. Além disso, essas propriedades também representam os dados brutos analisados de forma reduzida, que podem ser vídeos, medições de sensores, entre outros. Neste trabalho, busca-se evidenciar o potencial das propriedades modais, extraídas durante o processo de análise modal, como versões comprimidas do conjunto de medições original, especificamente em medições de vídeos com estruturas sob vibração. Para isso, um *framework* que realiza a extração das propriedades modais em vídeos de estruturas sob efeitos operacionais, utilizando técnicas para redução de dimensionalidade e extração de características, como a análise de componentes principais e separação cega de fontes, é implementado e adaptado para exportar as saídas resultantes dos seus procedimentos, como coordenadas modais, plano de fundo estático e modos de vibração. Além disso, alguns dos codecs mais utilizados na literatura, como o AVC/H.264, H.265/HEVC, VP9 e AV1 foram selecionados e implementados em um *pipeline* para obter resultados que pudessem ser comparados com os do *framework* desenvolvido. Para isso, também foram escolhidas algumas métricas para medir a qualidade dos vídeos, como a PSNR, MS-SSIM e VMAF, sendo essas essenciais para conseguir os dados que serão analisados. Após realizar as implementações dos *codecs*, métricas e *framework* de análise modal, uma análise comparativa é executada para alcançar o objetivo principal deste estudo. Assim, os dados obtidos revelam que, de fato, há um potencial da análise modal para a compressão de medições de vídeo. Para uma alta taxa de compressão fixa (todas as implementações foram analisadas para uma mesma taxa de compressão), a qualidade dos vídeos comprimidos pelo *framework* e pelos diversos *codecs* selecionados da literatura são equivalentes em métricas que levam em consideração o sistema de visão humano como um fator decisivo na percepção da qualidade em vídeos, como o MS-SSIM e o VMAF. Além disso, ainda em relação a essas mesmas métricas, os resultados provenientes do *framework* foram consistentemente melhores que os do *codec* H.264/AVC.

Palavras-chave: Análise modal, compressão de vídeos, structural health monitoring, extração de características.

ABSTRACT

In the area of structural health monitoring, modal analysis is used to verify the physical condition of structures, by extracting relevant features to damage identification. To achieve this, modal properties (natural frequencies, vibration modes, damping ratios) are generally utilized, because changes in the structures physical characteristics can be seen in its modal properties. Furthermore, these properties also represent the brute data, which can be videos, sensor measurements, and others, in a reduced form. In this work, we explore the potential of modal properties, extracted during modal analysis, as compressed versions of the original measurements, specifically in videos with vibrating structures. To achieve this, a framework that extracts modal properties from videos with vibrating structures by applying dimension reduction and feature extraction techniques, like principal component analysis and blind source separation, is implemented and adapted to export the result of its procedures (modal coordinates, static background and complex vibration modes). Furthermore, some of the most used codecs in the literature, H.264/AVC, H.265/HEVC, VP9 and AV1, were selected and implemented in a pipeline to gather data that could be compared with those collected from the developed framework. With that in mind, some video quality metrics, like PSNR, MS-SSIM and VMAF, were selected, those being essential to collect the data that will be analyzed. After implementing those codecs, metrics and the modal analysis framework, a comparative analysis is executed to achieve this paper's principal objective. The collected data shows that, in fact, modal analysis does have the potential to compress video measurements. The results describe that, for the same compression rate (all implementations were analysed in the same compression rate), the quality of videos compressed using the framework are equivalent to those compressed with the selected codecs of the literature in metrics that use the Human Vision System (HVS) as a decisive factor in the video quality perception, as MS-SSIM and VMAF. And for those same metrics, the results of the *framework* were consistently better than those of the codec H.264/AVC.

Keywords: Modal analysis, video compression, structural health monitoring, feature extraction.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Novas tecnologias para SHM e suas características. Fonte: Acervo do autor.	20
Figura 2 –	Linha do tempo representando as datas de lançamento de alguns dos principais <i>codecs</i> da literatura. Fonte: Acervo do autor.	23
Figura 3 –	Arquitetura do <i>codec</i> H.264/AVC Fonte: Acervo do autor.	24
Figura 4 –	Projeção dos dados em diferentes eixos ortogonais, gerando os autovetores. Fonte: (MEDEIROS; COSTA, 2008)	32
Figura 5 –	Em (a), alguns <i>frames</i> pertencentes ao vídeo da estrutura de três andares vibrando. Em (b), os <i>frames</i> são referentes a uma barra vibrando. Fonte: (YANG et al., 2017).	34
Figura 6 –	Medição que possui três barras vibrando em ângulos diferentes. Fonte: (SILVA; GREEN; MASCAREÑAS, 2021).	35
Figura 7 –	Processo de codificação representado de forma geral. Fonte: Acervo do autor.	35
Figura 8 –	Resultados esperados para o <i>framework</i> de análise modal. Fonte: Acervo do autor.	37
Figura 9 –	Coordenadas modais, PSD e fases extraídas para cada modo de vibração identificado no vídeo da barra, utilizando o <i>framework</i> proposto. Fonte: Acervo do autor.	40
Figura 10 –	Coordenadas modais, PSD e fases extraídas para cada modo de vibração identificado no vídeo da estrutura de três andares, utilizando o <i>framework</i> proposto. Fonte: Acervo do autor.	41
Figura 11 –	Coordenadas modais, PSD e fases extraídas para cada modo de vibração identificado no vídeo das três barras, utilizando o <i>framework</i> proposto. Fonte: Acervo do autor.	42

LISTA DE TABELAS

Tabela 1 – Frequências (em Hz) onde os modos de vibração foram identificados no vídeo da barra, a comparação foi feita entre a implementação de (YANG et al., 2017) e a utilizada neste trabalho. Fonte: Acervo do autor.	40
Tabela 2 – Frequências (em Hz) onde os modos de vibração foram identificados no vídeo da estrutura de três andares, a comparação foi feita entre a implementação de (SILVA; GREEN; MASCAREÑAS, 2021) e a utilizada neste trabalho. Fonte: Acervo do autor.	40
Tabela 3 – Frequências (em Hz) onde os modos de vibração foram identificados no vídeo das três barras, a comparação foi feita entre a implementação de (SILVA; GREEN; MASCAREÑAS, 2021) e a utilizada neste trabalho. Fonte: Acervo do autor.	41
Tabela 4 – Resultados obtidos para compressões feitas no vídeo da barra, que possui no total 27100KB. Fonte: Acervo do autor.	43
Tabela 5 – Resultados obtidos para compressões feitas no vídeo da estrutura de três andares, que possui no total 21100KB. Fonte: Acervo do autor.	44
Tabela 6 – Resultados obtidos para compressões feitas no vídeo das três barras, que possui no total 183000KB. Fonte: Acervo do autor.	44

LISTA DE ABREVIATURAS E SIGLAS

SHM	Structural Health Monitoring
BSS	Blind Source Separation
HVS	Human Vision System
AVC	Advanced Video Coding
HEVC	High-Efficiency Video Coding
PCA	Principal Component Analysis
BPP	Bits Per Pixel
PSNR	Peak signal-to-noise ratio
MS-SSIM	Multiscale Structural Similarity
VMAF	Video Multimethod Assessment Fusion
AOM	Alliance for Open Media
DCT	Discrete Cosine Transform

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Motivação	15
1.2	Justificativa	15
1.3	Objetivos	17
1.3.1	Objetivo Geral	17
1.3.2	Objetivos Específicos	17
1.4	Metodologia	17
2	REFERENCIAL TEÓRICO	19
2.1	Monitoramento da Condição Estrutural (SHM)	19
2.2	Codecs padrões na literatura	22
2.2.1	H.264/AVC (<i>Advanced Video Coding</i>)	24
2.2.2	H.265/HEVC (<i>High Efficiency Video Coding</i>)	25
2.2.3	VP8	25
2.2.4	AV1	26
2.3	Métricas para avaliar a compressão de vídeos	27
2.3.1	PSNR (<i>Peak Signal-To-Noise Ratio</i>)	28
2.3.2	MS-SSIM (Multiscale Structural Similarity)	29
2.3.3	VMAF (Video Multimethod Assessment Fusion)	30
2.4	Procedimentos para extrair características modais	31
2.4.1	Transformada de Hilbert	31
2.4.2	Análise de Componentes Principais	32
2.4.3	Separação Cega de Fontes	33
3	FRAMEWORK PARA COMPRESSÃO DE VÍDEOS UTILIZANDO ANÁLISE MODAL	34
3.1	Coleta de vídeos para o experimento	34
3.2	Visão geral do <i>framework</i> proposto	35
3.3	Módulo para armazenamento das propriedades modais	37
4	RESULTADOS	38
4.1	Implementação e configuração dos <i>codecs</i>	38
4.2	Extração das propriedades modais	38
4.3	Análise comparativa	42
5	CONCLUSÃO	45
5.1	Trabalhos futuros	45
	REFERÊNCIAS	46

1 INTRODUÇÃO

Os modos de vibração, e outras características modais, vêm sendo utilizados como parâmetros de extrema importância na área de monitoramento da integridade estrutural (do inglês, *Structural Health Monitoring* - SHM), para identificar a condição de estruturas. Isso se deve ao fato de que características modais (frequências naturais, modos de vibração e amortecimentos modais) extraídas de uma estrutura conseguem refletir fielmente as mudanças que ocorrem em relação às propriedades físicas dela (massa, amortecimento e rigidez), que por sua vez, são essenciais para identificar danos (FAN; QIAO, 2010). Porém, há uma característica pertencente as propriedades modais de estruturas que, apesar de conhecida, ainda não foi devidamente explorado pela literatura, a forma como os modos de vibração, frequências naturais e amortecimentos modais, informações comumente obtidas a partir da análise modal, são formas comprimidas do conjunto de dados original. Esse tipo de compressão é feita pois os valores modais são, de certa forma, visualizados e interpretados com mais facilidade por modelos matemáticos padrões (DOEBLING; FARRAR; PRIME, 1998), além de possuir os benefícios de uma compressão normal, atrelados diretamente ao volume de dados que serão processados.

Agora, há diversas formas para se extrair essas propriedades, e a mais utilizada nos dias atuais é a que faz uso de sensores acoplados fisicamente à estrutura, possibilitando a coleta dessas medições em pontos específicos dela. Porém, essa abordagem possui diversos problemas, como a necessidade de espalhar diversos sensores pela estrutura (aumentando o custo e a mão de obra necessária para o monitoramento) para que se possam obter dados que descrevem a estrutura com uma maior sensibilidade espacial, pois caso isso não seja possível, danos críticos localizados em partes não monitoradas podem gerar resultados catastróficos (YANG et al., 2017). Além disso, utilizar sensores em estruturas pequenas pode adicionar massa a elas, resultando em grandes chances de modificar o resultado final da medição (SILVA et al., 2020).

Levando essas questões em consideração, há claramente uma vantagem presente em utilizar abordagens que não exijam contato direto com a estrutura. Nesse sentido, foi observada a existência de trabalhos sendo feitos que pontuam a possibilidade de realizar a análise modal sem que haja a necessidade de sensores físicos, porém, algumas dessas abordagens ainda possuem um custo elevado em relação ao equipamento (como o monitoramento utilizando medidores de vibração a laser). Ou a necessidade de preparação prévia da superfície da estrutura, quando se utilizam técnicas de correlação de imagens digitais, em conjunto com câmeras digitais de baixo custo, que apesar de possuírem essa vantagem, necessitam que pontos sejam desenhados pela estrutura, para que algoritmos de rastreamento de pontos possam identificar as mudanças nela, o que será um obstáculo caso a estrutura seja muito grande, ou inacessível (YANG et al., 2017). Dito isso, uma forma de extrair as características modais de estruturas sem que haja a necessidade de preparação prévia da superfície, ou um custo elevado em relação aos equipamentos e manutenção se torna um fator importante. Felizmente, há algum tempo surgiram

algumas técnicas que, aliadas com a magnificação de vídeo, são capazes de obter as propriedades modais de estruturas através da extração de informações relacionadas aos movimentos contidos nas medições em vídeo, gravadas por câmeras de alta resolução e velocidade (CHEN et al., 2015).

No geral, os vídeos possuem dois tipos de informação: a informação espacial, contida em cada uma das imagens que compõem o vídeo, que são basicamente as informações relacionadas à luminosidade. E a informação temporal, correspondente ao desenvolvimento sequencial das informações contidas na imagem, que irá formar o vídeo em si no final. As imagens podem ser decompostas através de filtros, para obtenção de sinais de fase a amplitude que variam com o tempo, formatos de dados cuja manipulação é mais fácil que *pixels* brutos. A magnificação de vídeo é uma técnica que busca justamente conseguir uma representação do vídeo em forma de sinais, contendo informações dos deslocamentos dos *pixels*, para que as mesmas possam ser filtradas e amplificadas. E, ao final desse processo, o resultado será o vídeo reconstruído com os movimentos presentes nele intensificados (WADHWA et al., 2013). A ideia por trás de utilizar essa técnica para realizar a análise modal, é que a vibração pode ser representada pela movimentação dos *pixels*, então, ao extrair e analisar essa informação, é possível obter as propriedades modais referentes à estrutura.

Então, neste trabalho, um *framework* que extrai as características modais presentes em medições de vídeos de estruturas vibrando, utilizando técnicas para magnificação de vídeo e análise modal, contido no trabalho de (SILVA; GREEN; MASCAREÑAS, 2021) será adaptado, descrito e utilizado como base para verificar o potencial dessa abordagem em relação a compressão de vídeo, através de uma análise comparativa entre seus resultados e os de diferentes compressores disponíveis. Além disso, essa comparação será feita com base em métricas que são amplamente utilizadas para medir a qualidade e eficácia de compressões de vídeo.

Para fins de facilitar a exploração deste documento por parte do leitor, a sua estrutura foi definida da seguinte forma: capítulo 1, possuindo as seções de motivação, justificativa, objetivos e metodologia, para que fique claro a razão para existência deste trabalho e suas possíveis contribuições, além de expor quais passos foram necessários para o desenvolvimento da pesquisa; capítulo 2, necessário para contextualizar e descrever com detalhes as técnicas, ferramentas e áreas que são abordadas neste trabalho, além de familiarizar o leitor para com os conceitos utilizados; *framework* para compressão de vídeos utilizando análise modal, capítulo destinado a explicação aprofundada das etapas pertencentes ao *framework* abordado e utilizado nos testes; resultados, onde a análise comparativa será feita, e os dados gerados a partir da execução dos *codecs* e *framework* serão expostos, e além disso, será a seção do documento onde a implementação dos *codecs* será brevemente explicada, e os resultados referentes ao *framework* implementado serão validados; conclusão, último capítulo onde haverá uma breve discussão em relação aos resultados, dificuldades e possíveis trabalhos que poderão ser feitos ainda na área que tange tanto a compressão de vídeos, quanto a análise modal.

1.1 Motivação

Na área do monitoramento da integridade estrutural, o volume de dados sempre foi um fator problemático. Isso se deve grandemente à característica do monitoramento da estrutura que necessita de uma coleta de dados feita, em sua maioria, por um tempo mais longo, principalmente quando estamos falando de sensores físicos (HEO et al., 2018). Agora, é evidente que a medida que essas novas técnicas utilizando medições de vídeo vão emergindo, irão surgir diferentes empecilhos atrelados ao tipo de dado que está sendo utilizado. A princípio, sabemos que vídeos correspondentes a estruturas vibrando são compostos por, geralmente, centenas de milhares até milhões de *pixels*, além de possuírem de centenas a milhares de *frames* (MARTINEZ et al., 2020). Ou seja, se vídeos comumente utilizados na rede já exigem uma quantidade boa de largura de banda para serem transmitidos, e espaço em disco para serem armazenados, essas medições feitas para analisar as condições de estruturas, que devem ser de boa qualidade para serem processadas e analisadas, poderão se mostrar problemáticas devido ao seu tamanho.

Nesse sentido, é imperativo buscar formas de mitigar esse problema. Porém, note que estamos lidando com medições em vídeo, isso significa que há uma enorme literatura já existente, com diferentes técnicas para compressão desse tipo de dado, o que já resolveria boa parte do problema. Mas, a questão levantada nesse trabalho é que, como dito em (DOEBLING; FARRAR; PRIME, 1998), as técnicas presentes na extração de características modais já atuam como compressores dos dados originais. Isso abre algumas possibilidades como, reutilizar os recursos computacionais que já estão sendo investidos na análise modal para, de forma eficiente, também realizar a compressão das medições.

Então, o principal fator motivacional para o desenvolvimento deste trabalho seria o de explorar a possibilidade de que a compressão realizada em cima de dados relacionados a estruturas vibrando, através das diversas etapas pertencentes no processo de análise modal, seja algo relevante tanto para a área de monitoramento da integridade estrutural, quanto para a literatura relacionada a compressão em si, facilitando a transmissão e armazenamento de vídeos.

1.2 Justificativa

Quando analisamos estruturas com o auxílio de sensores físicos, podemos obter, por exemplo, 1024 pontos (uma para cada sensor). Levando em consideração que esses pontos serão obtidas diversas vezes ao decorrer do tempo (característica inerente do monitoramento) onde digamos que foram realizadas 100 medições, totalizando 102.400 pontos no total, a compressão desse volume de dados, utilizando as técnicas de análise modal e visão computacional pode ser expressada pela obtenção das características modais, como modos de vibração, amortecimentos modais, entre outros. Nesse caso em específico, caso 20 modos de vibração finais fossem obtidos, eles produziriam um conjunto de dados contendo apenas 2040 pontos representando os 20 modos (cada um possuindo sua representação em todas as 100 medições), 1 frequência de ressonância,

1 valor de amortecimento modal e 100 valores de amplitudes modais (DOEBLING; FARRAR; PRIME, 1998).

Tudo isso pode ser utilizado com sucesso para descrever as propriedades modais obtidas por sensores, e como elas conseguem representar os dados comprimidos. Agora, em relação ao objeto de estudo deste trabalho, as medições de vídeos, temos uma situação semelhante. Para exemplificar, imagine que um vídeo de uma estrutura vibrando possui 600 *frames*, e sua composição de *pixels* está disposta em uma matriz 216 x 384, totalizando 82.944 *pixels*. Note que, quando lidamos com esse tipo de medição, cada um dos *pixels* será considerado como um ponto que deverá ser efetivamente analisado pelos algoritmos. Se realizarmos a extração das propriedades modais diretamente atreladas a estrutura contidas nesses dados, mas dessa vez obtendo, por exemplo, apenas 6 modos, teremos no final um conjunto de dados totalizando 3.600 pontos referentes às coordenadas modais (uma matriz de ordem 600x6 com os valores de amplitude dos modos em relação ao tempo) e uma matriz de ordem 82944x6, expondo visualmente os modos de vibração, utilizada nas técnicas de magnificação de vídeo para reconstruir os dados originais a partir das propriedades modais.

Essa descrição mostra o potencial inicial dessa técnica para a compressão dos dados, note que o conjunto do exemplo foi reduzido de 49.766.044 pontos (representando todos os *pixels* presentes nos *frames* do vídeo, correspondente a 216x384x600), para apenas 3.600 atrelados às coordenadas modais e 497.644 (82.944x6) representando os modos de vibração, totalizando 501.244, uma compressão de aproximadamente 99% do conjunto de dados original. Agora, apesar desses resultados serem frequentes para esse tipo de técnica, não se vê na literatura os *frameworks* de análise modal sendo atrelados diretamente à compressão de dados, e essa situação se agrava mais quando falamos de análise modal utilizando medições de vídeo, visto que esta é uma abordagem de surgimento recente. Em sua maioria, as técnicas padrões para compressão de vídeos utilizam as regiões redundantes (geralmente estáticas) nos *frames* dos vídeos para realizar a codificação dos dados em blocos, ou buscam analisar *frames* anteriores ou posteriores para aplicar uma compensação de movimento, buscando prever a próxima posição dos *pixels* contidos nos *frames* do vídeo (GIROD, 2000).

Dito isso, atesta-se que apesar do potencial presente nos algoritmos pertencentes aos *frameworks* de análise modal para compressão de vídeos, não há um estudo que busque validá-lo através da comparação com *codecs* atuais. Além disso, a mudança no espectro dos dados para medições de vídeo exige algumas abordagens adicionais para gerenciamento do volume de dados, e a própria natureza das técnicas para extração de características pode ser utilizada como forte aliada nesse propósito. Esses fatores justificam uma avaliação da eficácia e qualidade das compressões feitas por um *framework* de análise modal em medições de vídeos contendo estruturas vibrando, por meio de uma análise comparativa com *codecs* padrões da literatura.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é realizar a implementação, adaptação e análise do *framework* contido no trabalho de (SILVA; GREEN; MASCAREÑAS, 2021), para a compressão de vídeos contendo estruturas vibrando.

1.3.2 Objetivos Específicos

- Realizar a implementação do *framework* de análise modal e magnificação de vídeo presente no trabalho de (SILVA; GREEN; MASCAREÑAS, 2021);
- Adaptar o *framework* para que seja possível armazenar os dados comprimidos em um formato que torne viável a análise comparativa;
- Realizar a implementação de alguns *codecs* de vídeos presentes na literatura (H.264, H.265, VP8);
- Implementar mecanismos para obtenção das métricas atreladas a qualidade de compressões de vídeos (PSNR, MS-SSIM, VMAF, Taxa de Compressão);
- Coletar resultados entre as diferentes implementações (*Framework* e *codecs*) e realizar a análise comparativa, utilizando três vídeos diferentes de estruturas vibrando, coletados de trabalhos que estudam a extração de características modais através de vídeos.
- Avaliar os resultados e definir vantagens e desvantagens em utilizar o *framework* para compressão de vídeos.

1.4 Metodologia

Primeiramente, um estudo aprofundado quanto às técnicas utilizadas para extração de características na área de SHM foi feita, e as principais ferramentas foram exploradas. Após isso, uma pesquisa mais específica quanto a *frameworks* que utilizam a magnificação de vídeo para auxiliar na análise modal em medições de vídeos foi feita, onde foram vistos alguns procedimentos matemáticos imprescindíveis para a execução dessas tarefa, representados por algoritmos para redução de dimensionalidade, extração de características através da separação de fontes, entre outros. Esses estudos foram essenciais para que a implementação do *framework* utilizado nesse trabalho fosse obtido e implementado.

A implementação do *framework* para análise modal através da visão computacional contido em (SILVA; GREEN; MASCAREÑAS, 2021) foi feita utilizando a linguagem de programação *python*, utilizando algumas bibliotecas para tratamento de sinais e vídeo. A parametrização

para que a análise modal necessita de uma análise quanto a primeira saída do *framework* como um todo, para reconhecimento da quantidade de modos de vibração da estrutura contida nos dados, esses poucos parâmetros serão descritos no capítulo 3

A seguir, uma pesquisa abrangente sobre a literatura de compressão de vídeos foi realizada, onde os principais *codecs* (H.264/AVC, H.265/HEVC e VP8) e métricas (PSNR, MS-SSIM e FMAF) foram obtidas. Essa etapa foi essencial para a implementação dos mesmos, e resultou em um segundo *framework*, também utilizando *python*, que reuniu um grupo selecionado de compressores de vídeo. Esse *framework* fez uso do programa *FFmpeg*, uma coleção de bibliotecas e ferramentas para processar conteúdos multimídia como áudio, vídeo, legendas, entre outros. Nesse caso, a parametrização se resumiu em estabelecer o BPP (*Bits Per Pixel*) referente ao vídeo após a compressão e descompressão, pois essa característica consegue influenciar diretamente a taxa de compressão e as métricas referentes a qualidade, que buscam comparar o quanto o vídeo comprimido foi distorcido em relação ao original. Esse parâmetro possibilita que o vídeo de saída dos compressores da literatura possuam uma taxa de compressão similar a do *framework* de análise modal, e isso é necessário visto que o *framework* não possui flexibilidade para modificar esse valor, significando que a análise comparativa vista no capítulo 4 buscará evidenciar a qualidade do vídeo comprimido para todas as implementações, utilizando taxas de compressão aproximadas.

Por fim, para a verificação das métricas em relação aos vídeos de saída dos *codecs* e *framework*, se utilizou o programa *MSU Video Quality Measurement*, que reúne diversas métricas presentes na literatura, e permite aplicá-las diretamente para comparação entre os vídeos originais e resultantes dos processos de compressão.

2 REFERENCIAL TEÓRICO

Nas seções referentes à esse capítulo, serão abordados em detalhes alguns assuntos relevantes para esse trabalho. Na primeira seção, as técnicas e características pertencentes a área de SHM serão evidenciadas, para que sua importância seja observada, tanto para este trabalho, quanto para a sociedade como um todo. Nas duas seções seguintes, haverá uma análise quanto a literatura relacionada a compressão de vídeos, onde os *codecs* e métricas utilizadas nos testes posteriores serão descritos. Por fim, na última seção, algumas técnicas matemáticas utilizadas no *framework* para análise modal utilizado nesse trabalho serão descritas, para que fique claro como sua manifestação nos algoritmos empregados impactam o processo de extração das características modais.

2.1 Monitoramento da Condição Estrutural (SHM)

No geral, o termo monitoramento da condição estrutural se refere ao desenvolvimento e implementação de estratégias utilizadas para detecção de danos, especialmente aplicadas em infraestruturas pertencentes às áreas de engenharia mecânica, civil e aeroespacial. Nesse caso, o conceito de dano é de extrema importância para que possamos entender as particularidades dessa área, e podemos dizer que há dois significados relevantes: o geral, onde é dito que dano pode ser definido como sendo toda e qualquer mudança introduzida em um sistema que cause problemas em sua performance atual ou futura; e o mais específico, referente a mudanças nas propriedades materiais e geométricas do sistema apenas (FARRAR; WORDEN, 2006). O conceito mais específico é importante pois limita o significado de dano, para que ele esteja mais próximo daquilo que será lidado na área de SHM.

Estruturas grandes como construções, pontes, turbinas e sistemas de dutos estão constantemente sob efeito de implicações causadas pelo ambiente, como terremotos, ventos, e temperaturas baixas ou altas. Além disso, a ação humana em si pode também desencadear mudanças na composição das estruturas citadas (SONY; LAVENTURE; SADHU, 2019). O dano causado em uma construção pode ser local, causado apenas em um ponto específico da estrutura, ou global, podendo ser detectado ou até mesmo visto por toda sua composição. De toda forma, as consequências esperadas para esse tipo de problema podem envolver perdas significativas relacionadas ao valor monetário daquela estrutura, ou equivalente ao tempo de inatividade e perda de eficiência dela devido aos danos. Além disso, em situações mais extremas, pode gerar um risco para vidas humanas.

Quando a condição estrutural de uma construção chega a um ponto onde manter suas propriedades físicas se torna insustentável, ocorre a falha. Falhas podem gerar consequências desastrosas, como por exemplo, quando uma ponte onde há a travessia constante de pessoas simplesmente é corrompida pelo dano estrutural, ou quando uma casa ou prédio desmorona.

Essas situações podem ser causadas por danos constantes e acumulados ao decorrer de um longo período de tempo, ou por eventos pontuais e de grande magnitude, gerando a falha em pouquíssimo tempo. Para ambas as situações, as técnicas envolvidas na área de SHM oferecem ferramentas de diagnóstico que possibilitam a detecção desses danos, até mesmo em tempo real (FARRAR; WORDEN, 2006).

Agora, o método utilizado para a identificação de danos através de SHM pode variar, mas a forma mais clássica seria a de obter medições relacionadas a alguma propriedade da estrutura (geralmente vibração), extrair características que possam descrever a situação atual dela, e analisar estatisticamente a sua condição. As medições são obtidas por sensores com ou sem fio que estão estrategicamente posicionados na superfície da estrutura, gerando dados discretos e normalmente esparsos. Essa abordagem, essencialmente, possui diversos problemas atrelados tanto ao custo para realizar a preparação do monitoramento, que envolve a obtenção e posicionamento estratégico dos sensores na estrutura, quanto para a manutenção, pois os sensores posicionados precisarão continuar funcionando para que o monitoramento possa ser feito durante um longo período, isso inclui proteger os sensores dos eventos que ocorrem no ambiente, manter sensores sem fio carregados, etc.

Um apanhado das principais tecnologias emergentes e algumas características que serão discutidas neste trabalho podem ser vistas na Figura 1. Recentemente, surgiram diversos estudos que exploram outras possibilidades para obtenção de dados para realizar a análise da condição de estruturas. No geral, essas tecnologias envolvem o uso de sensores presentes em *smartphones*, câmeras de alta velocidade, GPS, câmeras digitais, serviços em nuvem etc.

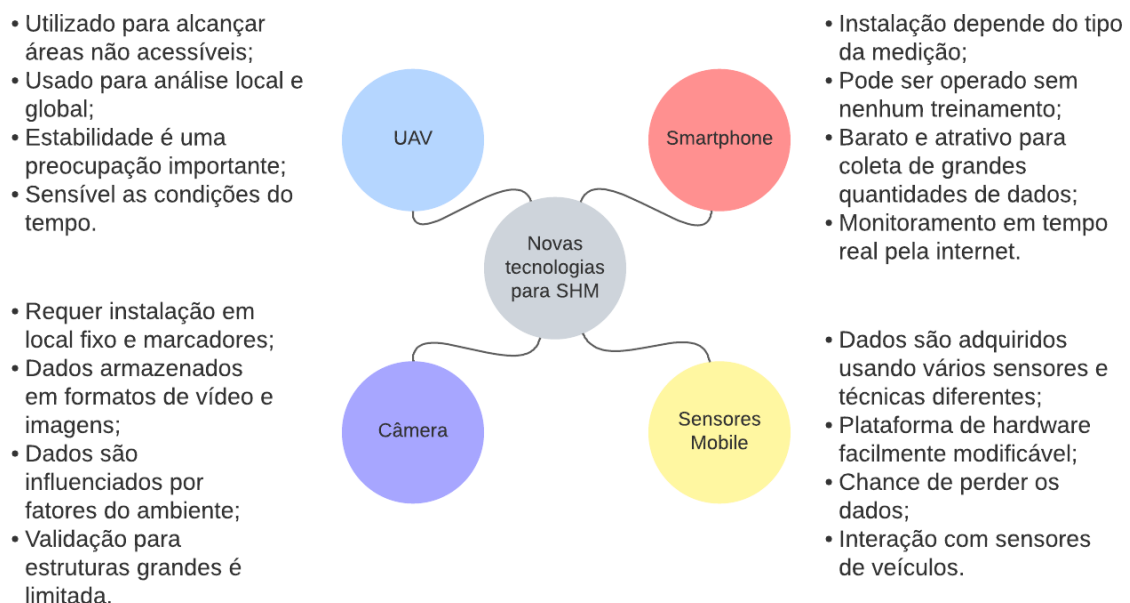


Figura 1 – Novas tecnologias para SHM e suas características. Fonte: Acervo do autor.

Devido a demanda para técnicas de detecção de danos para SHM utilizando sensores

que não necessitam de contato com a estrutura, houve um crescimento na utilização de UAVs (comumente chamados de *drones*), equipamentos que são muito úteis devido a sua mobilidade e flexibilidade para alcançar locais inacessíveis. Os *drones* podem ser utilizados para coleta de imagens enquanto voam ao redor da estrutura, e são controlados por um especialista que está no chão. Essas ferramentas conseguem coletar medições que podem ser utilizadas para obter informações de dano tanto global quanto local, devido a sua flexibilidade no momento da obtenção dos dados. Porém, necessitam de estabilidade no momento do voo para que a qualidade dos dados não seja prejudicada, indicando que as condições climáticas devem ser analisadas corretamente para que possam ser utilizados.

Agora, em relação aos *smartphones*. Desde 2010, há uma crescente popularização do uso de sensores presentes em *smartphones* para aplicações de SHM (SONY; LAVENTURE; SADHU, 2019). Isso ocorre pois esses dispositivos são baratos, móveis, possuem uma boa capacidade de armazenamento e um certo poder computacional, além disso, há sensores como acelerômetros nativamente atrelados a eles. Mas, há algumas limitações em relação ao uso de *smartphones* para SHM, como a falta de instruções para obter medições que representem corretamente as características necessárias na detecção de danos, e a baixa resolução dos acelerômetros presentes nesse tipo de dispositivo.

Prosseguindo, temos os sensores móveis, ferramentas robustas que são consideradas como a evolução dos sensores sem fio. Os sensores móveis são resultado de um esforço conjunto e multidisciplinar entre diversas áreas como a engenharia mecânica, civil, elétrica e ciência da computação. Isso acontece pois a utilização desse conceito envolve muito a área da robótica, com o objetivo de automatizar a obtenção de medições. Nesse caso, os dados são coletados diretamente na estrutura por robôs equipados com sensores e controlados por sistemas muito bem definidos e preparados. Através deste paradigma, é possível obter medições de forma eficiente e automatizada, porém, o custo envolvido e a complexidade para implementação formam um conjunto de desafios evidentes.

Finalmente, temos a utilização de câmeras de alta velocidade para obter medições com boa resolução e sensibilidade espacial. A utilização das técnicas atreladas a esse paradigma serão exploradas diretamente neste trabalho, onde será possível visualizá-las na prática. A ideia por trás dessa abordagem é que as imagens ou vídeos coletados sejam analisados por diversas técnicas da visão computacional, para que as características relevantes na estrutura sejam obtidas e utilizadas para indicar o estado dela. Geralmente, são utilizadas técnicas como DIC (Digital Image Correlation) e MM (Motion Magnification) para realizar o processamento das medições, e em alguns casos há a necessidade de preparar a superfície da estrutura para coleta, devido aos algoritmos de detecção de movimentos. Porém, há diversas abordagens que possibilitam contornar esse problema, e realizar as medições sem a necessidade de preparar a estrutura. Agora, há algumas limitações presentes nessa opção, como a forte influência de eventos externos nas medições obtidas, como o clima. Diferenciar modificações nas características da estrutura

causadas por danos ou eventos externos, como a chuva, é um desafio presente na utilização dessas câmeras.

2.2 Codecs padrões na literatura

A transmissão de vídeos com o passar do tempo vai exigindo cada vez mais largura de banda e espaço de armazenamento, devido ao crescente aumento da qualidade e volume desses dados. Além disso, nos dias atuais, a aplicabilidade de vídeos se expandiu grandemente, eles são utilizados em chamadas de vídeos, Internet das Coisas, Realidade Aumentada, jogos, etc. Por esses motivos, os *codecs*, que já vem sendo desenvolvidos há um bom tempo, necessitam acompanhar o constante crescimento do volume e qualidade de dados referentes aos vídeos. O termo codec significa *code-decoder*, e essas ferramentas são definidas como sendo um hardware, software ou uma combinação dos dois que aplica algoritmos de compressão em vídeos (PUNCHIHEWA; BAILEY, 2020). No geral, um codec vai definir e aplicar as regras presentes no ato da compressão e descompressão do vídeo original, e essa compressão modifica o formato do vídeo para um que os reprodutores de vídeo suportem, ou que um *decoder* padrão consiga decodificar.

Por várias décadas, o mercado de *codecs* foi dominado pelos grupos ISO/IEC e ITU-T, responsáveis pela criação, manutenção e desenvolvimento constante de diversos *codecs* como MPEG-2/H.262, H.263, MPEG-4 Visual e AVC (*Advanced Video Coding*), sendo AVC o codec mais comum e acessível. Posteriormente, os grupos MPEG e o VCEG estabeleceram a JCT-VC (*Joint Collaborative Team on Video Coding*) (GROIS et al., 2013), para que em 2013 pudessem desenvolver o novo H.265, também chamado de HEVC (*High Efficiency Video Coding*) (VANNE et al., 2012). O codec HEVC superou drasticamente o seu predecessor H.264/AVC, onde dependendo da configuração feita nos dois, o HEVC conseguiria obter uma redução de 40%-60% do *bitrate* total do vídeo em relação ao AVC, mantendo a mesma qualidade (LAUDE et al., 2019).

Uma visão geral de toda essa linha temporal, contendo alguns dos codecs citados, pode ser vista na Figura 2. Mais recentemente, temos VP8, VP9, Daala e Thor, codecs desenvolvidos por empresas como a Amazon, Facebook, Intel, Microsoft, Netflix, entre outras. Essas empresas, inclusive, formaram a AOM (*Alliance for Open Media*), que culminou no desenvolvimento do AV1, em 2018. Essa organização é concorrente direta do grupo JVET (*Joint for Video Exploration Team*), composta pelas duas organizações responsáveis pelos codecs padrões comentados anteriormente (AVC, HEVC, etc.), a ISO/IEC e ITU-T. Que se uniram novamente em 2015, e através de diversos estudos, definiram que um novo processo de padronização poderia ser feito, gerando consequentemente um novo codec, o VVC (*Versatile Video Coding*).

No geral, as tecnologias apresentadas aqui conseguiram evoluir de forma bastante acelerada, e esse avanço se beneficia principalmente do melhoramento de diversas técnicas per-

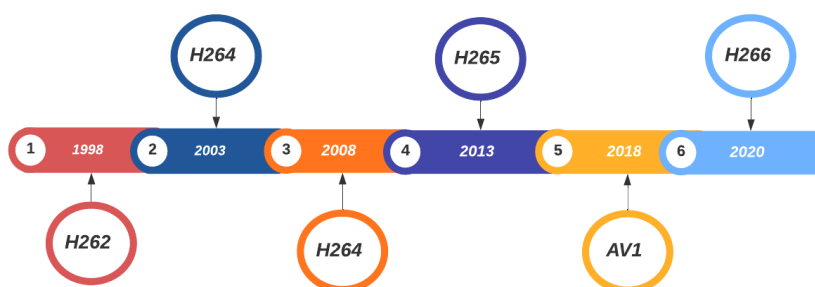


Figura 2 – Linha do tempo representando as datas de lançamento de alguns dos principais *codecs* da literatura. Fonte: Acervo do autor.

tencentos a outras áreas, como a de inteligência artificial (PUNCHIHEWA; BAILEY, 2020). Agora, durante o desenvolvimento de todas elas, o desafio principal da compressão se manteve o mesmo: reduzir o tamanho do vídeo transmitido ou armazenado o máximo possível, mantendo a qualidade dele o mais próximo da original. E as técnicas envolvidas nesse propósito fazem uso principalmente da exploração, e eliminação, de redundâncias presentes nos dados (RIPPEL et al., 2018).

Há dois tipos de redundâncias presentes nos vídeos que podem ser exploradas pelos algoritmos dos *codecs*: a temporal, redundância que pode ser encontrada ao comparar os diversos *frames* do vídeo entre si; e a espacial, presente apenas em um único *frame*. E na maioria dos casos, os algoritmos de compressão fazem uso de ambas, porém, há diversas formas de explorar cada uma dessas redundâncias. As técnicas que envolvem a análise dos diversos *frames* presentes no vídeo, para verificação das redundâncias contidas no *frame* atual são chamadas de *inter-frame*, e classificam os *frames* como *p-frames*, aqueles que são comprimidos através da análise de *frames* anteriores, e *b-frames*, aqueles que fazem uso de *frames* anteriores e posteriores para poderem ser comprimidos, geralmente este último possibilitando uma compreensão mais eficaz, porém, mais demorada e inflexível. Por fim, quando falamos de compressão que envolve a análise apenas do *frame* que será comprimido, como se fosse uma imagem, temos a existência dos *i-frames*, e esses são comprimidos utilizando algoritmos para codificação de imagens (WU; SINGHAL; KRÄHENBÜHL, 2018).

Dito isso, uma outra característica importante presente nas técnicas de compressão, é a possibilidade de compressão com perdas ou sem perdas. As compressões sem perdas permitem a codificação do vídeo mantendo a qualidade dele intacta, tornando o processo de compressão completamente reversível, porém, há uma limitação inerente a esse tipo de abordagem, diretamente atrelada à capacidade que ela tem de reduzir o tamanho do vídeo. Esse problema não existe quando falamos de compressão com perdas, onde o codificador irá reduzir o tamanho o máximo possível, inclusive sacrificando alguns dados presentes no vídeo, explorando a percepção humana para tornar os problemas gerados pela perda menos notáveis, o problema desta abordagem é que, a cada compressão, uma quantidade dos dados é perdida.

A seguir, o contexto histórico e os conceitos definidos serão aproveitados, para que seja

possível explorar com mais detalhes as peculiaridades presentes nos *codecs* apresentados. Note que, apesar de vários terem sido citados, apenas os que serão efetivamente implementados e utilizados na análise comparativa serão evidenciados aqui, as informações quanto a cada *codec* foram, em sua maioria, obtidas do trabalho de (ZHANG; MAO, 2019).

2.2.1 H.264/AVC (*Advanced Video Coding*)

O H.264/AVC (*Advanced Video Coding*) é sem dúvidas um dos *codecs* mais populares até os dias de hoje, muito adotado em diversas aplicações e plataformas, como *Youtube*, *Netflix*, *Samsung VR*, entre outros. Em sua arquitetura, há algumas etapas gerais que são utilizadas por diversos outros *codecs*, são elas: a codificação preditiva, a codificação por transformação, e a codificação por entropia. Todas essas etapas podem ser vistas na Figura 3, e serão descritas posteriormente.

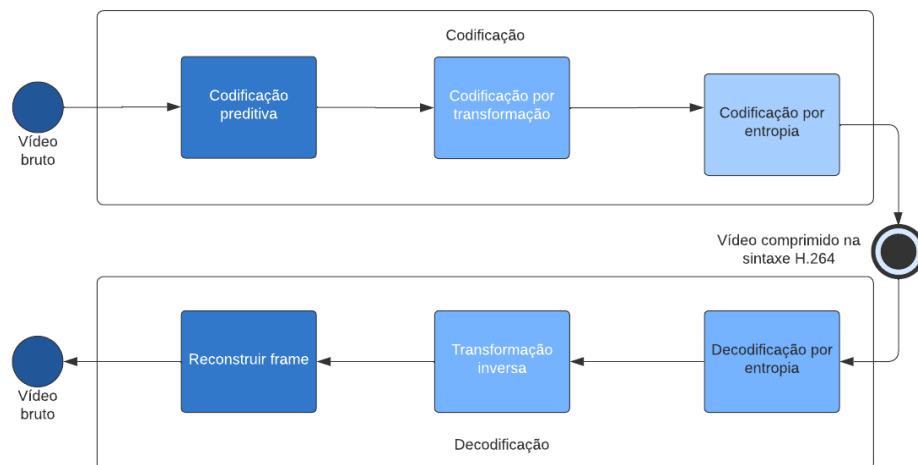


Figura 3 – Arquitetura do codec H.264/AVC Fonte: Acervo do autor.

Como sabemos, as codificações feitas utilizando *codecs* como o H.264/AVC fazem uso de blocos, que são basicamente divisões feitas nos *frames* pertencentes ao vídeo original, para que possam ser analisados meticulosamente. No H.264/AVC não é diferente, ele divide os *frames* em macro blocos de *pixels* que podem possuir tamanho de 4x4 até 16x16. Além disso, esse *codec* possuirá, assim como os outros, duas camadas principais, a de codificação (*video coding layer*), responsável por codificar o vídeo de fato em *bit streams*, e a de rede (*network abstraction layer*), que fará as definições quanto a cabeçalhos e outras informações necessárias para transmissão dos dados.

Agora, quanto aos passos presentes na camada de codificação, temos a codificação preditiva. Ela consiste em uma forma de obter informações quanto as redundâncias e movimentação dos *pixels* em uma determinada região (macro-bloco) do *frame* em relação aos outros. A codificação preditiva pode acontecer de duas formas: através da estimativa de movimento, buscando encontrar regiões nos *frames* passados ou futuros que coincidam com o *frame* analisado; ou

utilizando a compensação de movimento, onde a diferença entre a região analisada do *frame* atual e a mesma região nos demais *frames* é coletada (ZHANG; MAO, 2019). Esse processo descreve a primeira etapa para codificação, e os resultados dessas subtrações (resíduos) são armazenados para serem utilizados em sequência.

A segunda etapa, chamada de codificação por transformação, se refere a utilizar, por exemplo, uma transformada discreta de cosseno nos resíduos obtidos de cada macro bloco dos *frames* para obter coeficientes da DCT (*Discrete Cosine Transform*), esses coeficientes conseguem descrever as informações redundantes presentes nos blocos, e são essencialmente uma forma comprimida dessa representação.

Então, a última etapa é a transformação por entropia, e é utilizada para codificar os coeficientes e outras informações obtidas em *bit streams*, que é um tipo de informação muito conveniente para transmissão e armazenamento. Então, para obter a versão decodificada do vídeo, os coeficientes são primeiramente extraídos dos *bit streams* e utilizados na transformação inversa, para que seja possível obter de volta as informações residuais dos macro blocos, possibilitando a reconstrução dos *frames*.

Por fim, apesar de ainda ser muito utilizado, o *codec* H.264/AVC vem se tornando cada vez mais insuficiente para lidar com as demandas de vídeos atuais. Há diversas limitações, como a resolução máxima em que ele consegue codificar vídeos, que é 4K, o que não é suficiente, pois o 6K e 8K já vem se tornando qualidades utilizadas pelos clientes das plataformas de transmissão de vídeos.

2.2.2 H.265/HEVC (*High Efficiency Video Coding*)

O funcionamento do *codec* H.265/HEVC, em relação ao seu antecessor, não possui muita diferença. Pois, as etapas pertencentes ao processo de codificação são, de fato, as mesmas: Codificação preditiva, codificação por transformação e codificação por entropia. Porém, temos algumas particularidades, a primeira seria em relação ao tamanho dos blocos, que nesse caso, se tornam unidades de árvore de codificação, e possuem tamanhos de 64x64, 32x32, ou 16x16. Além disso, temos a utilização da técnica CABAC (*context adaptive binary arithmetic coding*) responsável pela etapa de codificação por entropia.

Apesar de possuir uma compressão muito mais eficaz que seu antecessor, o *codec* HEVC não é utilizado de forma tão ampla pelas plataformas, pois o tempo necessário para decodificação da informação pode ser até 10 vezes maior que o do H.264/AVC, o que pode implicar em problemas de lentidão ao tentar utilizá-lo em serviços que necessitem de velocidade.

2.2.3 VP8

O *codec* VP8 foi desenvolvido como parte de um projeto pertencente ao *Google*, chamado de *WebM*. Este projeto buscava desenvolver um formato de mídia que fosse disponível para

qualquer usuário da internet. Mas, para desenvolver o VP8, era necessário respeitar alguns critérios previamente estabelecidos, baseados essencialmente em demandas da rede atreladas ao consumo da largura de banda, capacidades de hardware dos usuários da rede, e formatos de vídeo e imagem existentes. Era necessário que o *codec* pudesse operar em qualidades de vídeo amplas, buscando trabalhar de forma leve, para que fosse possível operar em qualquer dispositivo, e oferecer uma ampla variedade de opções de qualidade e formato de imagens.

Apesar de possuir as mesmas etapas em relação aos outros *codecs*, o VP8 traz consigo algumas particularidades interessantes. A primeira em relação a divisão dos *frames*, onde cada um é separado em um macro bloco 16x16 (representando as informações de iluminação), e dois macro blocos 8x8 (chromas dos *pixels*: U e V). Também temos o fato de que, na codificação por transformação, esses macro blocos são reduzidos a apenas 3 macro blocos residuais de 4x4 cada, isso é feito através da extração dos coeficientes de DC (*discrete cosine*) de partes dos blocos pertencentes aos *frames*. Ou seja, com relação especificamente ao macro bloco 16x16, VP8 extrai os coeficientes residuais de cada um dos 16 blocos 4x4 dele, formando um novo bloco 4x4 contendo apenas as informações residuais. Além disso, nesse bloco de 4x4 resultante, uma transformada *Walsh-Hadamard* ainda é feita, sendo descrita como uma transformação de segunda ordem, reduzindo ainda mais as redundâncias presentes no bloco original de 16x16.

Outra característica interessante em relação ao VP8 seria o *golden frame*. No geral, esse *codec* utiliza três *frames* de referência diferentes, esses frames são utilizados para tentar, de certa forma, guardar uma informação substancial do que está acontecendo no vídeo durante a compressão, para que o algoritmo possa tomar decisões baseando-se em outros frames, não apenas no que está sendo analisado. Isso implica objetivamente que esse tipo de recurso está apenas disponível em compressões *inter-frame*. Em compressores antigos, geralmente há apenas um ou dois *frames* de referência, o *frame* anterior e o posterior ao que está sendo analisado, porém quando falamos do VP8, temos como referência o último *frame* do vídeo, um *golden frame*, e um terceiro *frame* que pode ser gerado de diversas formas pelo próprio *codec*, como pela junção de diversos *frames*, por exemplo. O *golden frame*, por outro lado, seria um *frame* de referência do passado, selecionado pelo próprio *codec*, e pode ser utilizado por ele para realizar tomadas de decisões ao decorrer da compressão (BANKOSKI; WILKINS; XU, 2011).

2.2.4 AV1

Desenvolvido pela AOM (*Alliance for Open Media*), o *codec* AV1 é uma combinação de diversos *codecs* (VP9, Daala e Thor), e consegue uma redução de 30% a mais que seu predecessor VP9, com a mesma qualidade (HAN et al., 2021). Utilizando o conceito de super blocos, o AV1 consegue representar um *frame* com blocos de até 128x128 de tamanho, podendo decompor os mesmos em blocos de até 4x4 (CHEN et al., 2015). E de forma diferente aos *codecs* vistos anteriormente, utiliza até 7 *frames* de referência distintos, sendo que o *frame* que está sendo codificado pode escolher qualquer um desses sete para utilizar nos algoritmos de predição.

Além disso, AV1 oferece diversos tipos de opções para compressões *intra-frame* e *inter-frame*. Para *intra-frames*, temos: A *directional intraprediction*, herdada de seu antecessor VP9; a *nondirectional smooth intraprediction*, *recursive intraprediction*, *chroma from luma prediction*, *intra-block copy* e *color palette*. E para predições inter-frames, AV1 disponibiliza as opções de *translation motion compensation*, *affine modal parameters* e *affine motion compensation*, *compound predictions*. Uma descrição mais detalhada de cada um desses modos, assim como diversas características adicionais presentes nesse codec em específico podem ser vistas no trabalho de (HAN et al., 2021).

2.3 Métricas para avaliar a compressão de vídeos

A utilização de técnicas para medir a qualidade de vídeos tende a ser utilizada, e possui a sua importância, principalmente em aplicações de videoconferência, video sob demanda e transmissão de vídeos (CARDOSO et al., 2012). Porém, com a aparição e desenvolvimento das diversas técnicas para compressão de vídeos, e a necessidade crescente de melhorar essas técnicas mais e mais, devido a alta demanda para transmissão e armazenamento de vídeos, se tornou imperativo também utilizá-las para avaliar a eficácia de *codecs*, com a finalidade de compará-los. Essa comparação é essencial para validar novas abordagens e decidir se uma determinada técnica superou, de alguma forma, as demais presentes na literatura.

No geral, há dois tipos de métricas para medição da qualidade de vídeos, as que utilizam testes subjetivos, que são mais adequadas, porém, necessitam de mais tempo e introduzem uma maior complexidade, tornando sua utilização inviável em aplicações de tempo real, e as de teste objetivo. Porém, a avaliação da qualidade de vídeos geralmente é feita por testes subjetivos, pois o conceito de qualidade varia de um indivíduo para o outro, isso acontece pois diferentes tipos de usuários vão possuir necessidades distintas, consequentemente, uma qualidade de vídeo boa para um, pode ser ruim para o outro.

Mas, a avaliação subjetiva da qualidade necessita de um alto custo, tempo e de uma metodologia específica, que pode ser complexa. Então, as métricas objetivas surgiram justamente para resolver esse problema, e trazer um novo paradigma, contendo técnicas rápidas, imparciais, confiáveis e de baixo custo para a análise da qualidade de vídeos. Essas técnicas se baseiam fortemente no conhecimento de limites da percepção de degradações pelo HVS (Sistema Visual Humano), e utilizam esse fator como uma das principais propriedades para análise da qualidade do vídeo. A classificação das métricas objetivas pode ser feita de acordo com a disponibilidade de referências para comparação, e na maioria dos casos, estamos falando do vídeo original. Temos então: métricas que atuam na comparação direta entre o vídeo original (supondo que ele esteja em sua condição perfeita) e o degradado para obter a qualidade resultante, chamadas de métricas com referência total; métricas com referência reduzida, utilizando apenas algumas informações acerca do vídeo original para realizar a comparação e obter resultados; e as métricas

sem referência, que como o nome diz, não possuem acesso ao vídeo original, realizando a análise da qualidade apenas utilizando as informações contidas nos *frames* do vídeo analisado.

A seguir, serão descritas com mais detalhes algumas das métricas utilizadas na literatura para analisar a qualidade de vídeos, e que serão usadas posteriormente para obter os resultados avaliados na análise comparativa. Note que foram selecionadas principalmente métricas com referência total, pois o objetivo é justamente realizar uma comparação direta entre o vídeo original e o comprimido.

2.3.1 PSNR (*Peak Signal-To-Noise Ratio*)

A primeira métrica que será abordada neste trabalho é a PSNR (*Peak Signal to Noise Ratio*). Esta técnica faz parte do grupo que realiza uma medição sensível ao erro, indicando que a funcionalidade dela aborda justamente a extração do erro entre o vídeo original e o degradado, que nesse caso, se refere ao vídeo pós compressão, a fórmula geral para esse tipo de processo pode ser descrita por:

$$|Err| = |S - S_r|, \quad (2.1)$$

onde Err corresponde ao erro obtido, S aos frames do vídeo original e S_r representa os frames do vídeo degradado, a subtração dessas duas matrizes é feita, em módulo, para coletar a diferença entre elas, essa diferença (pixel por pixel) é justamente o erro. As fórmulas apresentadas aqui foram retiradas do trabalho de (CARDOSO et al., 2012), assim como algumas informações adicionais sobre algumas métricas objetivas. Prosseguindo, essas técnicas são utilizadas justamente pela sua simplicidade matemática e baixo custo computacional, a PSNR não é uma exceção, e para que seja implementada, a seguinte expressão deve ser seguida:

$$PSNR = 10 \log_{10} \left[\frac{255^2}{MSE} \right], \quad (2.2)$$

o segredo para obtenção dos erros que são utilizados na PSNR está contido na obtenção do MSE (*mean square error*), que pode ser calculado a partir do valor médio dos erros quadráticos entre os *pixels* do vídeo original e do vídeo degradado (CARDOSO et al., 2012), representado a seguir:

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - h_i)^2, \quad (2.3)$$

para obter o valor do MSE, basta computar a média atrelada a quantidade de amostras N, representada por $\frac{1}{N} \sum_{i=1}^N$. Após isso, a componente da equação $(f_i - h_i)^2$ adiciona o conceito de erro ao cálculo do MSE, isso significa que f_i e h_i , nesse contexto, seriam os *pixels* originais e

degradados, respectivamente. Além disso, essa última componente é elevada ao quadrado, para que valores positivos sejam obtidos.

2.3.2 MS-SSIM (Multiscale Structural Similarity)

SSIM (*single-scale structural similarity*) é conhecido de forma abrangente na literatura de métricas para medir a qualidade de vídeos. Sua primeira descrição foi feita no trabalho de (WANG; SIMONCELLI; BOVIK, 2003), e desde então, vem sendo modificada e utilizada de diversas formas. O SSIM admite que o sistema de visão humano (do inglês, *human vision system* - HVS) é altamente capaz de extrair informações estruturais de cenas, por isso, ele utiliza uma abordagem de similaridade estrutural entre a medição degradada e de referência para realizar a comparação. Sua descrição matemática acontece da seguinte forma: leve em consideração que os sinais que serão comparados podem ser descritos como $(x = x_i | i = 1, 2, \dots, N)$ e $(y = y_i | i = 1, 2, \dots, N)$, compostos por valores discretos e não-negativos, representando, por exemplo, dois *frames*. O SSIM operará extraindo, primeiramente, resultados correspondentes a média de x (μ_x), variância de x (σ_x^2) e covariância de x e y (σ_{xy}), pois a média e a variância conseguem expressar as informações de luminosidade e contraste de x , e a covariância consegue descrever a tendência dos valores de x e y de mudarem juntos, indicando uma similaridade estrutural.

A extração das informações de luminosidade (l), contraste (c) e similaridade estrutural (s), a partir da média, variância e covariância é feita da seguinte forma:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (2.4)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (2.5)$$

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (2.6)$$

onde C_1 , C_2 e C_3 são pequenas constantes definidas com base no intervalo dos valores dos *pixels* contidos nos *frames* do vídeo (geralmente de 0 até 255). Com base na obtenção desses valores que descrevem os dois objetos analisados, podemos definir, finalmente, a equação que representa o cálculo definitivo para o SSIM:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (2.7)$$

dessa forma, os resultados obtidos a partir dessa métrica, sendo limitados pelo próprio cálculo, vão até 1, sendo esse valor um máximo único, que acontecerá apenas quando x for exatamente igual a y (resultado perfeito).

Agora, por mais que o SSIM consiga descrever de uma forma satisfatória a similaridade estrutural entre dois *frames* em um vídeo, foram desenvolvidas diversas versões diferentes desta mesma técnica com o decorrer do tempo, abordagens essas que conseguem explorar não só as informações estruturais, e contraste e luminosidade para obter os resultados. Esse é o caso da métrica MS-SSIM (*multi-scale structural similarity*), que herda diversas características do SSIM, mas também analisa outras informações que afetam a forma como o usuário percebe a qualidade do vídeo, como a distância do plano do vídeo ao observador, e a densidade da amostra do sinal do vídeo, e claro, a capacidade de percepção do HVS do observador, informações que influenciam fortemente nos resultados dados por métricas subjetivas, que agora são também usados nessa métrica objetiva.

A partir do vídeo de referência e do vídeo distorcido, o sistema aplica um filtro de passa baixa e reduz a resolução espacial por um fator de 2 em cada iteração. A comparação de contraste $c_j(f, h)$ e a de similaridade estrutural $s_j(f, h)$ são calculadas até a j -ésima iteração, enquanto a comparação de luminância $l_M(f, h)$ é feita apenas na iteração de número M , ou seja, M define o tamanho da escala (CARDOSO et al., 2012). Sendo assim, a representação de todo o processo de avaliação pertencente ao MS-SSIM pode ser descrito como:

$$MS-SSIM(f, h) = [l_M(f, h)]^{\alpha M} \cdot \prod_{j=1}^M [c_j(f, h)]^{\beta j} [s_j(f, h)]^{\gamma j}, \quad (2.8)$$

considere aqui que os expoentes αM , βj e γj são constantes que podem ser modificadas para dar prioridade pra algum dos componentes presentes na equação, esta abordagem também é utilizada no SSIM, mas os autores da métrica descrevem em seu trabalho que preferem manter essas constantes como 1.

2.3.3 VMAF (Video Multimethod Assessment Fusion)

O VMAF (video multimethod assessment Fusion), desenvolvido pela *Netflix*, é uma métrica focada primariamente na medição da qualidade em vídeos que sofreram um *rescaling* ou foram comprimidos por algum *codec*. Portanto, sua utilização para realizar a análise comparativa neste trabalho é adequada. Seu funcionamento se traduz na utilização de aprendizado de máquina e a teoria por traz da modelagem da visão humana para criação de um modelo que seria capaz de estimar a percepção humana em relação a qualidade de vídeos. Para isso, ela coleta os resultados de diversos algoritmos para medição da qualidade, e funde eles utilizando um SVM (*support vector machine*) (RASSOOL, 2017). Esses algoritmos utilizados são nada mais que métricas que medem a qualidade de imagens, que são utilizadas para medir a qualidade dos vídeos *frame por frame*, algumas das métricas inseridas nesse processo são: AN-SNR (*anti-noise SNR*), DLM (*detail loss measure*), VIF (*visual information fidelity*) e MCPD (*mean co-located pixel difference*).

A computação dos resultados de todas essas métricas, e a sua combinação, gerarão no final o que chamamos de *VMAF score*, esse score vai de 0 a 100 para cada *frame* do vídeo, sendo 100 considerado o melhor resultado, quando o *frame* de referência é igual ao vídeo supostamente degradado (GARCÍA et al., 2019).

2.4 Procedimentos para extrair características modais

Agora, o esforço será voltado para o entendimento de quais técnicas e procedimentos matemáticos são geralmente utilizados para a extração das características modais. Na área de SHM, diversos passos são necessários para que seja possível identificar a condição atual da estrutura monitorada, um desses passos se refere a extração de características das medições coletadas sobre a estrutura. Essas características podem ser extraídas de diversas formas, mas no geral, teoremas matemáticos bem estabelecidos, e procedimentos para o processamento de sinais são considerados suficientemente adequados para essa tarefa.

A seguir, introduziremos algumas das principais técnicas utilizadas para a extração de características na análise modal, principalmente quando se tratam de medições de vídeos com estruturas vibrando. Esses mesmos procedimentos são utilizados efetivamente no *framework* para análise modal que será posteriormente descrito no próximo capítulo deste trabalho, onde uma visualização mais prática da utilização dessas técnicas poderá ser vista.

2.4.1 Transformada de Hilbert

A utilização da Transformada de Hilbert na área da análise de vibrações vem sendo constantemente estudada durante dezenas de anos (FELDMAN, 2011), isso se deve principalmente ao fato de que essa técnica aplicada a sinais consegue expor algumas informações quanto a amplitude, fase, e frequência de vibrações, que são características muito importantes, e são utilizadas amplamente na literatura para o processamento de sinais. Além disso, devido ao fato de que a transformada de Hilbert (do inglês, *Hilbert transform* - HT) possui a característica de extrair, principalmente, a frequência das vibrações, se tornou uma possibilidade que a operação inversa também fosse possível. Ou seja, através de uma frequência fornecida, com o uso da transformada de Hilbert seria possível a identificação da vibração contida no sinal.

Para que seja possível extrair as informações já descritas, que são utilizadas para a análise modal e extração de características em um sinal de entrada, a transformada de Hilbert converte a função desse sinal para uma analítica, que nesse caso, é essencialmente complexa, indicando que a saída desse procedimento será sempre uma função analítica complexa, com um componente real e outro imaginário. A equação que representa um sinal analítico representado com o auxílio da transformada de Hilbert pode ser visto a seguir, onde:

$$x_a(t) = x(t) + j\tilde{x}(t), \quad (2.9)$$

nesse caso, o sinal analítico, representado por $x_a(t)$ será obtido quando a própria função for atrelada ao resultado proveniente da transformada de Hilbert dela mesma, que seria $\tilde{x}$, representando juntamente com j a parte imaginária do par de Hilbert. Por fim, para que a própria HT seja computada, é necessário seguir a seguinte equação:

$$\tilde{x}(t) = H[x(t)] = x(t) * \frac{1}{\pi t} = \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t - \tau)} d\tau. \quad (2.10)$$

2.4.2 Análise de Componentes Principais

O PCA (*principal component analysis*) é dito como sendo uma das técnicas da estatística multivariável mais populares que existem, e é utilizada por quase todas as disciplinas científicas, além disso, é provavelmente a técnica mais antiga para análise estatística multivariável (ABDI; WILLIAMS, 2010). Esta técnica pode ser utilizada para analisar *datasets* de n dimensões, representados por matrizes contendo uma coluna representando as *features* presentes no conjunto de dados, e linhas onde é possível visualizar as observações correspondentes a cada *feature*.

No geral, os objetivos atrelados a execução do PCA se resumem a extrair as informações mais importantes do conjunto de dados, comprimir a quantidade de dados mantendo apenas as informações mais importantes (característica muito importante para este trabalho), e analisar a estrutura das observações e das variáveis como um todo. Agora, para alcançar esses objetivos, o PCA obtém projeções do conjunto de dados original de uma forma que a variação contida no *dataset* seja propriamente representada neste novo eixo, sendo que a cada novo componente principal gerado, essa variação descrita vai sendo cada vez mais reduzida (MEDEIROS; COSTA, 2008). Um exemplo que pode mostrar mais claramente como esses novos eixos se parecem pode ser vista na Figura 4. Note que se os dados forem projetados no eixo correspondente ao componente principal 1, a variação descrita pelos dados será maximizada, enquanto o componente principal 2, que é ortogonal ao primeiro, irá representar uma quantidade reduzida da variação, caso os dados sejam projetados nele.

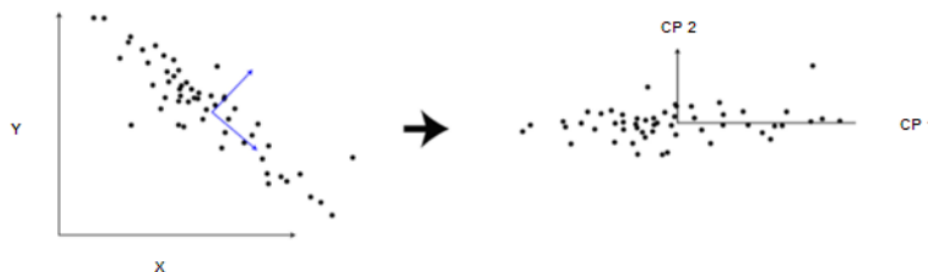


Figura 4 – Projeção dos dados em diferentes eixos ortogonais, gerando os autovetores. Fonte: (MEDEIROS; COSTA, 2008)

O procedimento necessário para computar o PCA inicia com um pré-processamento no conjunto de dados, com o objetivo de centralizá-lo em torno da origem, pois isso facilita

os cálculos seguintes. A seguir, ao obter a matriz X com os dados centralizados, é necessário encontrar a matriz de covariâncias empíricas de X :

$$\Sigma = \frac{1}{n} (XX^T). \quad (2.11)$$

A matriz de covariância também pode ser expressada utilizando o o teorema da decomposição espectral:

$$\Sigma = U \Lambda U^T, \quad (2.12)$$

nesse caso, U será uma matriz contendo os autovetores associados a matriz de covariância de X , e Λ uma matriz diagonal que possui os autovalores associados a cada um dos autovetores. Obter essas matrizes seria o mesmo que resolver o problema do PCA, pois os autovetores atrelados a matriz de covariância representarão a direção que mais maximiza a variação dos dados, consequentemente indicando os componentes principais, e os autovalores serão números que evidenciarão a importância de cada um dos autovetores, acusando quais representam mais essa variação.

2.4.3 Separação Cega de Fontes

A separação cega de fontes (do inglês, *Blind Source Separation* - BSS) é uma técnica promissora para o processamento de sinais de forma não supervisionada, e já vem sendo utilizado na área de estruturas dinâmicas há um tempo (YANG; NAGARAJIAH, 2013). O que faz o BSS ser uma técnica capaz de ser inserida na análise modal é a sua capacidade de extrair e recuperar sinais escondidos que estão misturados em um único sinal.

As técnicas que aplicam a identificação modal assumem que as medições que estão sendo analisadas são misturas lineares dos componentes que geram o sinal, sendo assim, com relação ao BSS, esse paradigma seria representado matematicamente da seguinte forma (MCNEILL; ZIMMERMAN, 2008):

$$x^{(mx1)} = A^{(mxn)} s^{(nx1)}, \quad (2.13)$$

na equação 2.13, a contante A representaria a matriz de mistura, que foi utilizada para embutir de forma linear as componentes do sinal, que são representadas aqui por s . O objetivo principal do BSS seria de estimar tanto a matriz de mistura A quanto os componentes separados em s . Há diversas abordagens utilizadas tanto na área de processamento de sinais, como na própria análise modal para aplicar o *blind source separation*, como o ICA (*independent component analysis*) e o CP (*complexity pursuit*).

3 FRAMEWORK PARA COMPRESSÃO DE VÍDEOS UTILIZANDO ANÁLISE MODAL

O objetivo deste trabalho é utilizar o *framework* para análise modal em medições de vídeos com estruturas vibrando, descrito no trabalho de (SILVA; GREEN; MASCAREÑAS, 2021), com o intuito de realizar uma análise em cima da compressão de vídeos a partir desse tipo de abordagem. Sendo assim, este capítulo é completamente dedicado a explicar como o *framework* foi implementado, levando em consideração o seu fluxo geral, as medições que serão analisadas por ele, e as técnicas que compõem cada etapa. Além disso, na última seção, a forma utilizada para armazenar as matrizes correspondentes as saídas da análise modal será evidenciada, além da motivação para realizar esse tipo de armazenamento.

3.1 Coleta de vídeos para o experimento

Para que os testes possam ser feitos, e o *framework* possa ser descrito com mais clareza, é necessário que as medições utilizadas sejam descritas. Neste trabalho, alguns vídeos de estruturas vibrando foram coletados de outros artigos que lidam com a análise modal aplicada a esse tipo de medição. Dito isso, temos na Figura 5 alguns *frames* que foram retirados de dois vídeos que serão utilizados para realização das análises: O vídeo (a), referente a uma estrutura de três andares vibrando, que possui 400 *frames* no total, e uma dimensão de 216x384 *pixels*; e o vídeo (b), que contém uma barra vibrando, 600 frames e dimensão exatamente igual ao vídeo da estrutura.

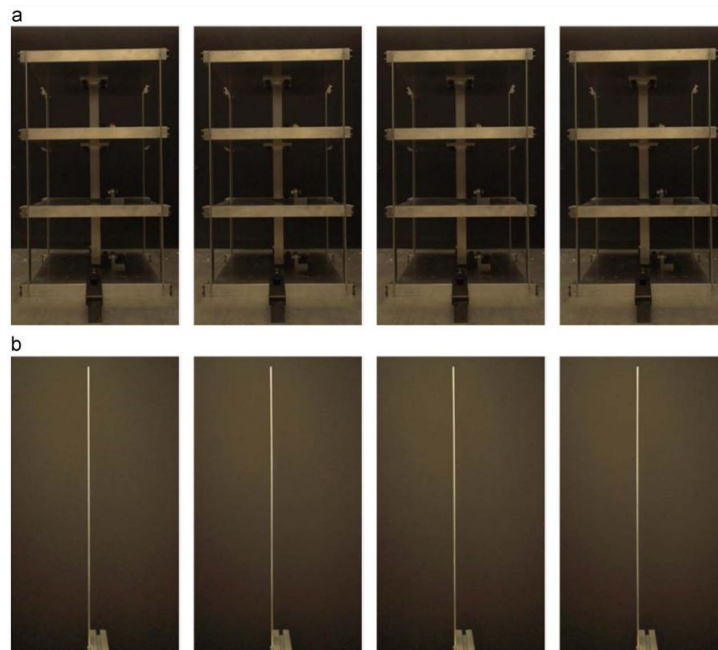


Figura 5 – Em (a), alguns *frames* pertencentes ao vídeo da estrutura de três andares vibrando. Em (b), os *frames* são referentes a uma barra vibrando. Fonte: (YANG et al., 2017).

Além dessas medições, também foi adquirido um vídeo que possui três barras diferentes

vibrando, que pode ser visualizado na Figura 6. Essa medição em específico apresenta um desafio maior para o *framework*, pois há uma quantidade elevada de modos de vibração diferentes que devem ser extraídas dele. Quanto a sua especificação, ele possui 1242 *frames*, e dimensão deles é de 544x332 *pixels*.

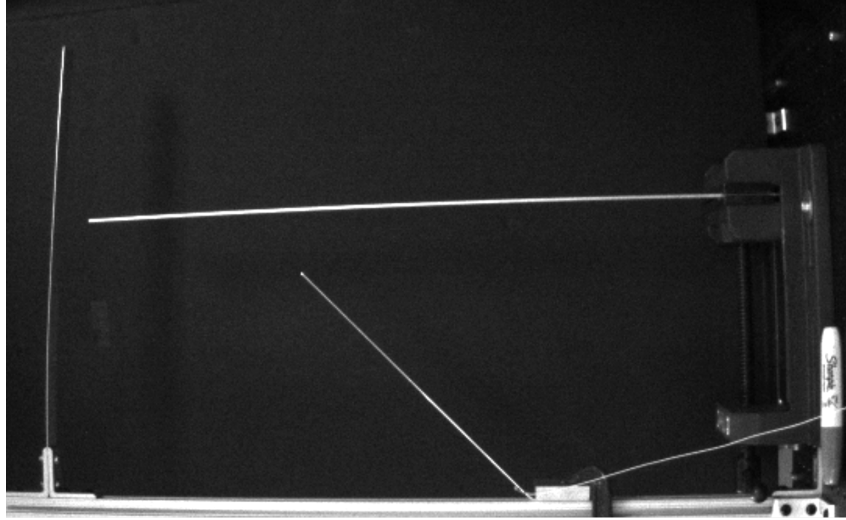


Figura 6 – Medição que possui três barras vibrando em ângulos diferentes. Fonte: (SILVA; GREEN; MASCAREÑAS, 2021).

3.2 Visão geral do *framework* proposto

As etapas pertencentes ao processo de extração de características modais do *framework* podem ser vistas na Figura 7, que também evidenciam as técnicas que atuam na compressão dos dados. Ao decorrer desta seção, essa estrutura será descrita com detalhes. Porém, o módulo para exportação das propriedades modais será apresentado apenas na próxima seção.

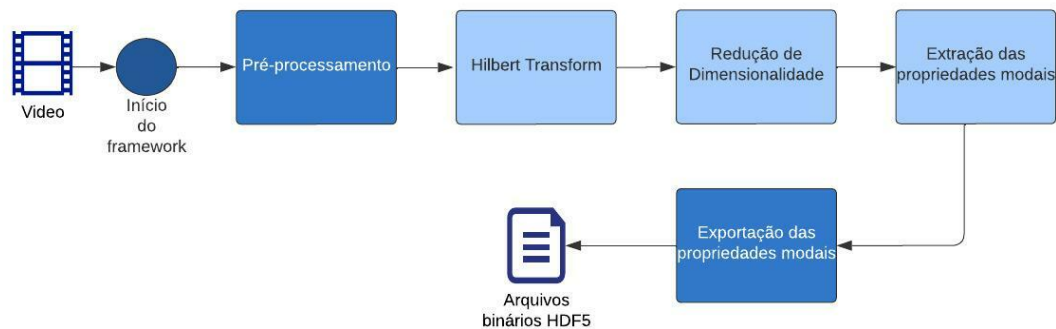


Figura 7 – Processo de codificação representado de forma geral. Fonte: Acervo do autor.

Inicialmente, o vídeo é carregado no *framework* e um pré-processamento inicial é feito, que se resume a centralização dos dados, e também na retirada do *background*, que é desnecessário para a análise modal, visto que as propriedades modais serão extraídas apenas da estrutura vibrando. Para que isso seja feito, uma característica presente em todos os vídeos que

são utilizados para validar a análise modal em medições nesse tipo de abordagem é explorada: o plano de fundo estático. O plano de fundo ser estático significa que, ao decorrer dos *frames* que compõem o vídeo, os *pixels* pertencentes ao plano de fundo não tem seus valores modificados, não há uma mudança notável nos valores desses *pixels*, então, ter a informação do plano de fundo de um único *frame* também significa possuir de todos os outros. A partir disso, com o vídeo representado por uma matriz X , contendo colunas representando cada pixel presente no vídeo, e as linhas o histórico temporal dos valores desse pixel ao decorrer dos *frames*, a remoção do *background* pode ser feita a partir da subtração da média dos valores dessa matriz nela mesma:

$$\tilde{X} = X - \text{Mean}(X), \quad (3.1)$$

nesse momento, o valor da média da matriz X é guardado, e será um componente importante no processo de reconstrução do vídeo comprimido, visto que toda a informação de *background* será armazenada nele. Note que esse procedimento pode ser considerado como uma “compressão”, porém, ele é bastante simples, e pode facilmente ser melhorado para desenvolver ainda mais o potencial do *framework* para compressão de vídeos. Após o pré-processamento, temos a aplicação das diversas técnicas para extração de características, cada uma com sua finalidade específica.

No caso da transformada de Hilbert, ela é aplicada as medições principalmente para obter uma versão complexa dos dados, com componentes reais e imaginárias, o que é muito importante para obter as informações de magnitude e fase em relação aos sinais contidos no vídeo de forma mais clara. Após isso, o PCA (*principal component analysis*) é aplicado nas saídas reais e imaginárias da transformada de Hilbert, com dois principais objetivos: reduzir o volume de dados, mantendo as informações que o descrevem, e iniciar uma extração de características, deixando os modos de vibração em evidência.

Por fim, a extração das propriedades modais é propriamente feita utilizando o BSS (*blind source separation*). Para isso, os resultados correspondentes aos componentes principais do PCA são multiplicados pela matriz de mistura que foi obtida através do BSS, dessa forma, é possível extrair diretamente as fontes que estão contidas no sinal. E com os resultados obtidos no BSS (fontes e matriz de mistura) e no PCA (componentes principais, autovetores e autovalores), é possível finalmente extrair as propriedades modais, representadas pelas coordenadas modais e os modos de vibração. Na Figura 8 é possível perceber de uma forma visual todas as saídas pertencentes as técnicas aplicadas.

Na componente (a) da Figura 8, temos uma demonstração de como os números correspondentes aos autovalores vão reduzindo, nesse caso, o autovetor correspondente ao primeiro autovalor (correspondente ao maior número) consegue representar melhor a variação contida no conjunto de dados. Agora, na componente (b), os componentes principais em relação ao domínio de frequência podem ser visualizados. Os componentes (c) e (d) são a matriz de mistura, e as

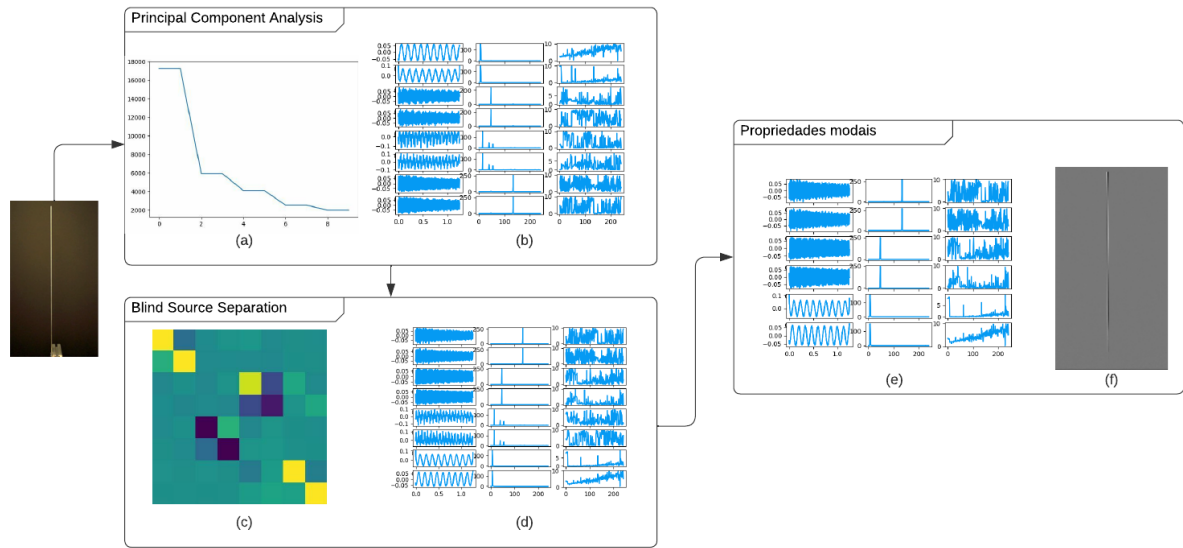


Figura 8 – Resultados esperados para o *framework* de análise modal. Fonte: Acervo do autor.

fontes, respectivamente. E os itens (e) e (f) são os resultados esperados para o *framework*, sendo eles as coordenadas modais e os modos de vibração, respectivamente. Os modos de vibração descrevem a forma da vibração, e as coordenadas modais representam a frequência e a amplitude dessa vibração.

3.3 Módulo para armazenamento das propriedades modais

O módulo de exportação das propriedades modais busca realizar o armazenamento dessas informações em um arquivo binário. Esse método foi escolhido devido a flexibilidade desse tipo de arquivo, e sua boa compressão com relação a medições em matrizes. Nos testes feitos neste trabalho, o módulo *H5PY*, do *python* foi utilizado para armazenar os resultados no formato binário *HDF5*.

Com isso, é possível exportar as coordenadas modais, os modos de vibração e o *background* retirado do vídeo de forma adequada, esses arquivos podem ser transmitidos e armazenados por um baixo custo, pois representam a versão comprimida do vídeo. Por fim, para que o vídeo seja reconstruído (etapa correspondente aos módulos de decodificação em compressores da literatura), é necessário que os arquivos binários sejam carregados no *framework*, e que as coordenadas modais sejam multiplicadas pelos modos de vibração, permitindo a obtenção de uma única matriz maior, contendo todos os frames do vídeo. Então, o *background* extraído no início deve ser somado a essa matriz, com o objetivo de devolver o plano de fundo estático.

4 RESULTADOS

Este capítulo é destinado unicamente a coleta e análise dos resultados provenientes da análise comparativa entre os *codecs* da literatura e o *framework* para análise modal. Porém, haverá uma seção destinada também para evidenciar como os *codecs* foram implementados e configurados para realização da compressão e coleta das métricas.

4.1 Implementação e configuração dos *codecs*

Os *codecs* utilizados efetivamente neste trabalho (H.264/AVC, H.265/HEVC, VP9 e AV1) foram implementados através de um *pipeline* que faz uso da ferramenta *FFMPEG*, um programa que reúne as métricas e *codecs* mais utilizados na literatura, e oferecem uma grande variedade de configurações e parâmetros que as execuções sejam flexíveis e adaptáveis para cada caso. Nesse sentido, como o *framework* para análise modal não é flexível em relação ao tamanho das propriedades modais, que são fixos, há uma limitação quanto a taxa de compressão, característica essa que pode ser relevada, visto que a taxa de compressão é geralmente bem alta.

Devido a essa limitação, os *codecs* foram configurados para gerar resultados com taxas de compressão aproximadas as do *framework*, significando que os resultados mais relevantes para essa análise serão as saídas das métricas, ou seja, a qualidade do vídeo comprimido em relação ao original. Para que essa configuração pudesse ser feita, o conceito de BPP foi implementado para definir um *bitrate* específico para os vídeos de saída das compressões, impactando diretamente na qualidade do vídeo de saída. A fórmula que foi utilizada para definir um *bitrate* específico, a partir de um BPP selecionado, pode ser vista a seguir:

$$bitrate = \frac{BPP \times m \times p \times fps}{1000}, \quad (4.1)$$

sendo m e p as dimensões de pixels referentes aos frames do vídeo. No mais, os *codecs* no geral foram configurados para rodarem na opção de maior eficiência, chamada de *slower*, indicando que a compressão será feita da melhor forma possível, sem considerar o tempo necessário para codificação. O tempo não é uma informação relevante para este trabalho, pois a análise modal é um procedimento que, normalmente, toma bastante tempo, significando que não é o objetivo desta técnica performar melhor que os *codecs* da literatura nesse sentido, vistos que estes são desenvolvidos e aperfeiçoados para atuarem em aplicações em tempo real, por demanda, etc.

4.2 Extração das propriedades modais

Para que a compressão dos vídeos seja feita através do *framework* de análise modal descrito anteriormente, é necessário utilizá-lo para extrair as propriedades modais das estruturas.

Essas propriedades se resumem aos modos de vibração, que indicam a forma da vibração feita pela estrutura, e as coordenadas modais, que vão indicar a amplitude daquele movimento, em relação ao tempo. Porém, existem algumas outras características atreladas ao processo de análise modal que podem ser averiguadas não para comprimir os dados, mas para que o sucesso da identificação dos modos possa ser confirmado.

A densidade espectral (do inglês, *power spectral density* - PSD) pode ser considerada como uma dessas características. Ela consegue descrever, no domínio de frequência (medido em Hertz), onde está localizado o modo de vibração em questão. Para obter a PSD, é necessário que a transformada de Fourier (nesse caso, utilizou-se a transformada rápida de Fourier, ou *fast fourier transform* - FFT) seja computada em cima das coordenadas modais X , o módulo elevado ao quadrado disso (pois a PSD é uma função real e positiva) seria a PSD. A equação que descreve esse procedimento pode ser vista logo abaixo.

$$PSD = |FFT(X)|^2, \quad (4.2)$$

outra informação importante que pode ser coletada através das coordenadas modais é a de fase. A fase é de extrema importância para os *frameworks* baseados em magnificação de vídeos, pois é nela que se encontram as informações quanto aos deslocamentos dos *pixels*. Ao acessar e amplificar a fase referente ao vídeo, é possível identificar e revelar até mesmo os pequenos movimentos que acontecem nele. Na Figura 9, é possível visualizar os primeiros resultados do *framework* utilizado neste trabalho, em relação ao vídeo da barra, as coordenadas modais, PSD e as informações de fase foram diretamente extraídas do vídeo através das técnicas para extração de características.

Note que na Figura 9, temos pares de coordenadas modais que são referentes aos modos reais (em azul) e imaginários (vermelhos), devido aos resultados referentes a transformada de *Hilbert*. Agora, na PSD, a frequência com a maior amplitude representa onde está contido o modo de vibração encontrado, essa informação (extraída neste trabalho) foi comparada diretamente com a obtida no trabalho de (YANG et al., 2017), e os resultados de ambos os estudos podem ser vistos na Tabela 1. O fato dos resultados coincidirem é de extrema importância para validar a análise modal executada, pois no estudo de (YANG et al., 2017) esses dados são comparados com os provenientes de sensores fisicamente acoplados às estruturas contidas nos vídeos, ou seja, também é uma validação para as técnicas de extração das propriedades modais em vídeos, em relação às tradicionais.

Como essas informações foram obtidas para todos os vídeos utilizados neste trabalho, na Figura 10 também temos figuras atreladas as coordenadas modais, PSD e fases obtidas no vídeo da estrutura de três andares.

Além disso, a mesma comparação entre as frequências encontradas para os três modos no vídeo da estrutura de três andares consta na Tabela 2.

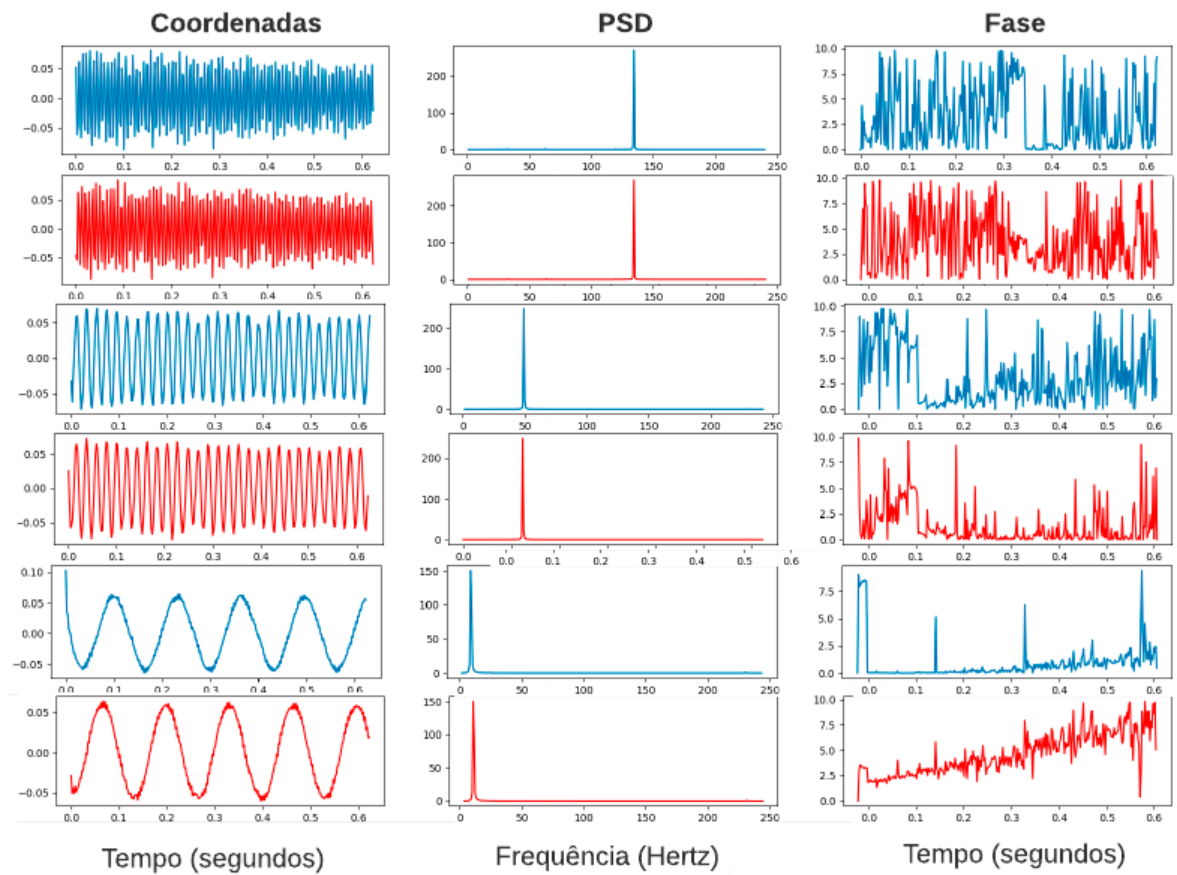


Figura 9 – Coordenadas modais, PSD e fases extraídas para cada modo de vibração identificado no vídeo da barra, utilizando o *framework* proposto. Fonte: Acervo do autor.

Modo	<i>Framework</i> implementado	<i>framework</i> de (YANG et al., 2017)
1	7.20	7.50
2	47.20	48.20
3	133.60	133.60

Tabela 1 – Frequências (em Hz) onde os modos de vibração foram identificados no vídeo da barra, a comparação foi feita entre a implementação de (YANG et al., 2017) e a utilizada neste trabalho. Fonte: Acervo do autor.

Modo	<i>Framework</i> implementado	<i>framework</i> de (YANG et al., 2017)
1	6.60	6.31
2	18.00	17.92
3	25.79	25.92

Tabela 2 – Frequências (em Hz) onde os modos de vibração foram identificados no vídeo da estrutura de três andares, a comparação foi feita entre a implementação de (SILVA; GREEN; MASCAREÑAS, 2021) e a utilizada neste trabalho. Fonte: Acervo do autor.

Por fim, temos as mesmas informações sendo obtidas para o vídeo das três barras na Figura 11, dessa vez, foram extraídos dois modos de vibração diferentes para cada barra contida no vídeo, resultando em 6 pares de modos no total.

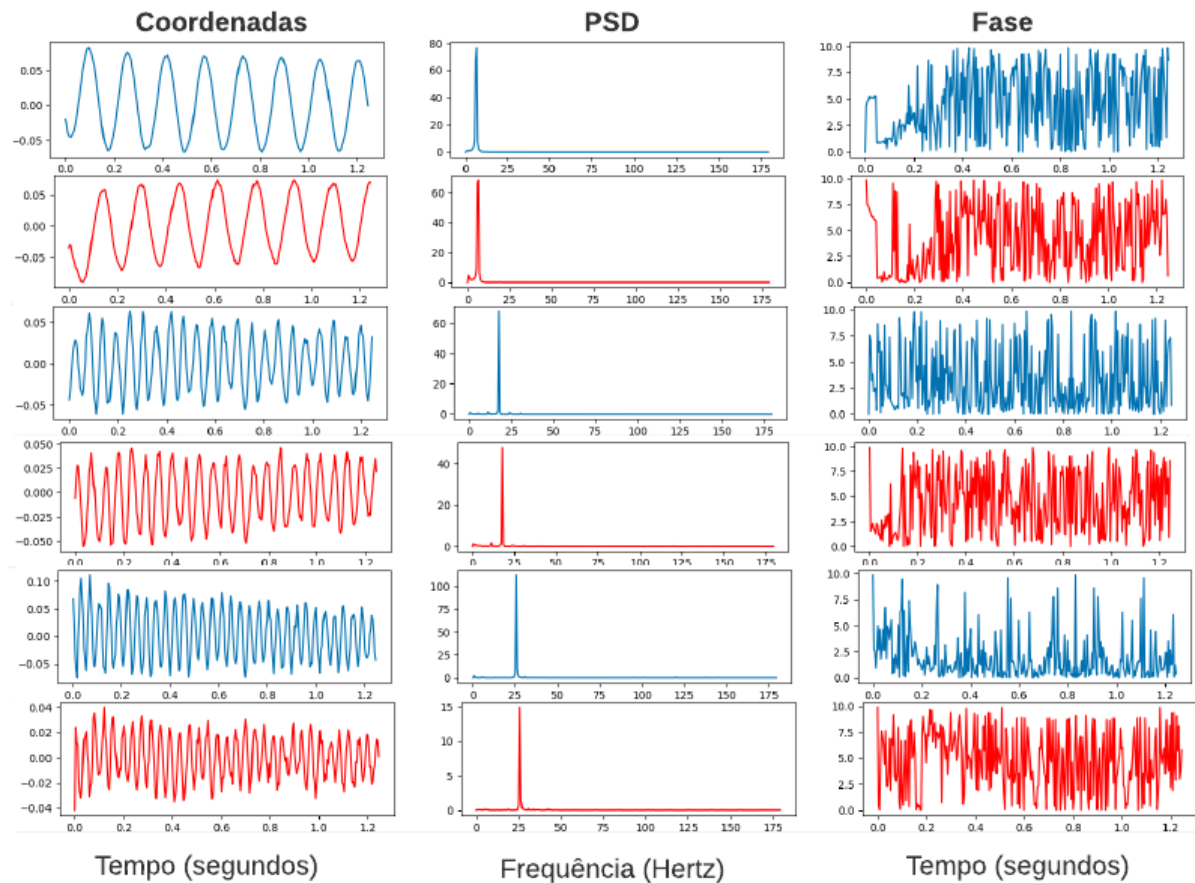


Figura 10 – Coordenadas modais, PSD e fases extraídas para cada modo de vibração identificado no vídeo da estrutura de três andares, utilizando o *framework* proposto. Fonte: Acervo do autor.

Neste caso, as frequências encontradas pelo *framework* para o vídeo das três barras foram comparadas com aquelas vistas no trabalho de (SILVA; GREEN; MASCAREÑAS, 2021), estudo onde o vídeo é analisado. Essa comparação pode ser vista na Tabela 3.

Modo	<i>Framework</i> implementado	<i>Framework</i> de (SILVA; GREEN; MASCAREÑAS, 2021)
1	7.32	7.26
2	44.56	44.42
3	6.42	6.40
4	41.13	41.00
5	11.14	11.10
6	22.28	22.21

Tabela 3 – Frequências (em Hz) onde os modos de vibração foram identificados no vídeo das três barras, a comparação foi feita entre a implementação de (SILVA; GREEN; MASCAREÑAS, 2021) e a utilizada neste trabalho. Fonte: Acervo do autor.

Visto que os resultados referentes as frequências em que os modos de vibração foram identificados coincidem com os encontrados em outros trabalhos para as mesmas medições, é possível validar a veracidade dessas propriedades modais, e utilizá-las efetivamente para alcançar o objetivo principal deste estudo.

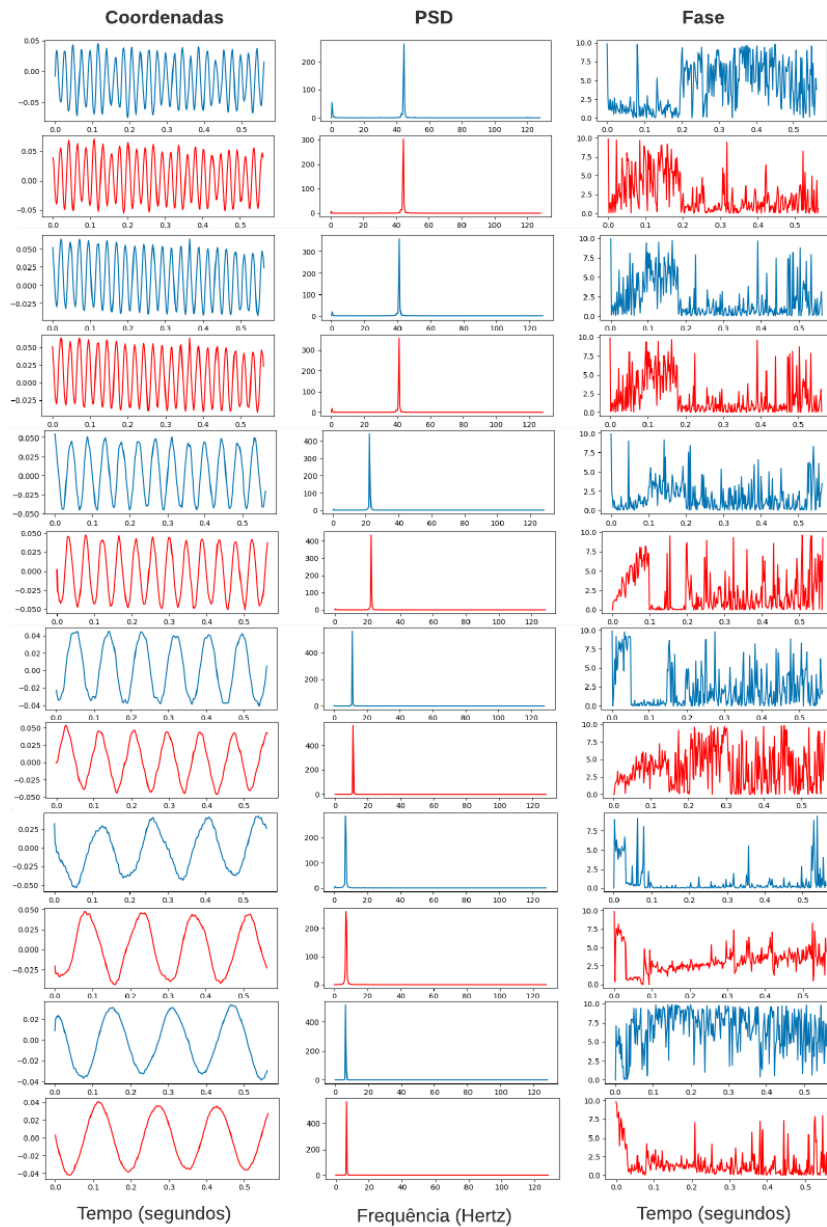


Figura 11 – Coordenadas modais, PSD e fases extraídas para cada modo de vibração identificado no vídeo das três barras, utilizando o *framework* proposto. Fonte: Acervo do autor.

4.3 Análise comparativa

Dito isso, os dados a seguir, contidos na Tabela 4, Tabela 5 e Tabela 6, mostram os resultados obtidos pelas métricas citadas neste trabalho, em relação aos vídeos comprimidos pelo *framework* e pelos principais *codecs* abordados. Note que em todos os resultados, um esforço foi feito para que a taxa de compressão pudesse ser aproximada em todas as técnicas abordadas.

É evidente que o *framework* de análise modal se aproxima de forma consistente dos *codecs* mais utilizados na literatura em relação as métricas de qualidade, para as taxas de compressão utilizadas. Em alguns casos, as saídas obtidas ao aplicar as métricas no *framework* são, inclusive, levemente melhores que as obtidas pelos *codecs*, isso acontece frequentemente

quando analisamos os resultados do MS-SSIM e VMAF. Esses dados são interessantes pois, no geral, a compressão por análise modal se sai melhor nas métricas que levam em consideração o sistema de visão humano do que na PSNR, que essencialmente só leva em consideração as informações atreladas ao erro entre o vídeo original e o comprimido.

Isso pode ser explicado pelo fato de as métricas sensíveis ao erro, como a própria PSNR, realizarem essa avaliação de uma forma bem direta, não levando em consideração informações importantes para o sistema de visão humano, como a informação estrutural. Adicionalmente, o processamento simples feito pelo *framework* para lidar com o plano de fundo estático também prejudica os resultados do dele em relação a PSNR, pois o *background* representa boa parte dos *frames*, e também contém muitas informações redundante, que são extremamente exploradas pelos *codecs*.

Agora, esses fatores são coerentes quando é levado em consideração que o *framework* foi feito para realizar análise modal, onde o plano de fundo estático é realmente irrelevante, e não necessariamente a compressão das medições. Porém, mesmo que esse não seja o foco para o *framework*, e ele extraía informações muito relevantes para a análise modal e SHM enquanto comprime os dados, ele conseguiu resultados equivalentes aos de diversos codecs. Por exemplo, na Tabela 5, em resultados obtidos pela métrica MS-SSIM, a compressão feita pelo *framework* conseguiu uma pontuação idêntica as dos outros compressores, até mesmo em relação ao AV1, que é considerado um dos mais novos e sofisticados *codecs*.

Além disso, o *framework* consegue consistentemente pontuações melhores que o H.264/AVC tanto no MS-SSIM, quanto no VMAF. Isso acontece em quase todas as situações, exceto na Tabela 6, onde a pontuação para o VMAF do H.264/AVC foi melhor que a do *framework*. No mais, é possível ver algumas situações onde a compressão através da análise modal supera, mesmo que levemente, *codecs* mais sofisticados como o VP9 e o H.265, como na Tabela 6, em relação a métrica MS-SSIM, onde o *framework* conseguiu pontuações melhores que todos os *codecs* exceto o AV1. E na Tabela 5, onde ele obteve resultado melhor para a métrica VMAF do que o H.265/HEVC.

Compressor	PSNR	MS-SSIM	VMAF	Taxa de compressão
Framework	45,30	0,996	93,67	99,36%
H.264	48,51	0,996	92,85	99,36%
H.265	51,94	0,997	94,75	99,34%
VP9	52,18	0,997	94,70	99,37%
AV1	52,72	0,997	95,18	99,57%

Tabela 4 – Resultados obtidos para compressões feitas no vídeo da barra, que possui no total 27100KB. Fonte: Acervo do autor.

Note, no entanto, que os *codecs* foram configurados para comprimir o vídeo utilizando uma alta taxa de compressão, e essa configuração geralmente não é utilizada em aplicações de tempo real. Isso também significa que para todos os compressores analisados, seria possível

Compressor	PSNR	MS-SSIM	VMAF	Taxa de compressão
Framework	47,39	0,998	95,80	98,40%
H.264	48,44	0,997	94,95	98,41%
H.265	50,08	0,998	95,71	98,36%
VP9	50,94	0,998	96,23	98,32%
AV1	51,41	0,998	96,69	99,51%

Tabela 5 – Resultados obtidos para compressões feitas no vídeo da estrutura de três andares, que possui no total 21100KB. Fonte: Acervo do autor.

Compressor	PSNR	MS-SSIM	VMAF	Taxa de compressão
Framework	41,42	0,991	95,31	99,26%
H.264	42,61	0,990	96,54	99,28%
H.265	42,89	0,990	95,18	99,27%
VP9	43,20	0,990	96,35	99,25%
AV1	45,47	0,994	97,74	93,99%

Tabela 6 – Resultados obtidos para compressões feitas no vídeo das três barras, que possui no total 183000KB. Fonte: Acervo do autor.

comprimir os vídeos com uma taxa de compressão menor, garantindo uma maior qualidade no resultado. Dito isso, é possível afirmar que no contexto de análise modal, é possível reaproveitar os recursos computacionais investidos na extração de propriedades modais para realizar a compressão das medições envolvidas nesse processo, facilitando a transmissão e armazenamento desses dados. É interessante pontuar também que, para além da compressão, também é possível utilizar os resultados que o *framework* gera para realizar o monitoramento da integridade estrutural da construção presente no vídeo. Além disso, esses dados também constataam que a qualidade do vídeo, mesmo após aplicar uma alta taxa de compressão, não é tão prejudicada, característica importante para viabilizar a reutilização das medições comprimidas.

Caso o leitor deseje visualizar a implementação do *framework* utilizado neste trabalho, essa está disponível publicamente, e pode ser acessada em um repositório no *Github*, em (SILVA, 2022).

5 CONCLUSÃO

Neste trabalho, uma análise comparativa é feita entre os *codecs* mais utilizados na literatura para codificação de vídeos, e um *framework* que extrai propriedades modais, para validar a sua utilização na compressão de vídeos referentes a estruturas vibrando, paradigma recente que vem sendo muito estudado nas áreas de visão computacional, SHM (*structural health monitoring*) e análise modal. Para isso, foram selecionadas diversas métricas objetivas para analisar a qualidade de vídeos que são muito relevantes para avaliar a eficácia de técnicas de compressão. Agora, a partir da análise dos dados coletados, foi possível evidenciar que há uma grande vantagem em usar o próprio *framework* de análise modal para comprimir e exportar as medições que são usadas nesse processo, através de propriedades modais como modos de vibração e coordenadas modais, visto que o volume de dados diminui de forma eficiente, e além disso, mantendo uma boa qualidade em relação ao vídeo original. Isso pôde ser visto com clareza na análise comparativa, onde as pontuações do *framework* se mantiveram equivalentes e até mesmo melhores que as dos *codecs* da literatura em métricas que levam em consideração informações importantes nos vídeos, como estruturais, distância da visualização, e o sistema de visão humano no geral. Além disso, também se constata um potencial que deve ser explorado futuramente em relação a esse tipo de abordagem, visto que esta não utiliza, necessariamente, as redundâncias contidas nos *frames* do vídeo para realizar a compressão, como os outros *codecs* disponíveis fazem, mas sim uma extração de características, que é capaz de expressar o próprio movimento contido no vídeo.

5.1 Trabalhos futuros

Como possíveis áreas de exploração dentro do *framework*, temos que melhorias poderiam ser certamente feitas na etapa de remoção do *background*, que apesar de ser algo secundário no processo de análise modal como um todo, é um procedimento que tem algum impacto quando falamos de sua utilização para a compressão das medições. Além disso, temos a questão da criptografia, que pode ser muito mais explorada e aplicada na área da segurança em sistemas que realizam SHM, para manter a confidencialidade das medições que contém informações sensíveis. No mais, investigações futuras podem ser feitas quanto a técnicas diferentes, e talvez mais eficientes, para realizar a análise modal, e consequentemente a compressão das medições.

REFERÊNCIAS

- ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, Wiley, v. 2, n. 4, p. 433–459, jun. 2010. Disponível em: <https://doi.org/10.1002/wics.101>.
- BANKOSKI, J.; WILKINS, P.; XU, Y. Technical overview of VP8, an open source video codec for the web. In: **2011 IEEE International Conference on Multimedia and Expo**. IEEE, 2011. Disponível em: <https://doi.org/10.1109/icme.2011.6012227>.
- CARDOSO, J. et al. Comparação das métricas objetivas de qualidade de vídeos baseadas na similaridade estrutural e na sensibilidade ao erro. **Revista de Tecnologia da Informação e Comunicação**, GN1 Genesis Network, v. 1, n. 2, p. 33–40, abr. 2012. Disponível em: <https://doi.org/10.12721/2237-5112.v01n02a05>.
- CHEN, J. G. et al. Modal identification of simple structures with high-speed video using motion magnification. **Journal of Sound and Vibration**, Elsevier BV, v. 345, p. 58–71, jun. 2015. Disponível em: <https://doi.org/10.1016/j.jsv.2015.01.024>.
- DOEBLING, S. W.; FARRAR, C. R.; PRIME, M. B. A summary review of vibration-based damage identification methods. **The Shock and Vibration Digest**, SAGE Publications, v. 30, n. 2, p. 91–105, mar. 1998. Disponível em: <https://doi.org/10.1177/058310249803000201>.
- FAN, W.; QIAO, P. Vibration-based damage identification methods: A review and comparative study. **Structural Health Monitoring**, SAGE Publications, v. 10, n. 1, p. 83–111, abr. 2010. Disponível em: <https://doi.org/10.1177/1475921710365419>.
- FARRAR, C. R.; WORDEN, K. An introduction to structural health monitoring. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 365, n. 1851, p. 303–315, dez. 2006. Disponível em: <https://doi.org/10.1098/rsta.2006.1928>.
- FELDMAN, M. Hilbert transform in vibration analysis. **Mechanical Systems and Signal Processing**, Elsevier BV, v. 25, n. 3, p. 735–802, abr. 2011. Disponível em: <https://doi.org/10.1016/j.ymssp.2010.07.018>.
- GARCÍA, B. et al. Practical evaluation of VMAF perceptual video quality for WebRTC applications. **Electronics**, MDPI AG, v. 8, n. 8, p. 854, jul. 2019. Disponível em: <https://doi.org/10.3390/electronics8080854>.
- GIROD, B. Efficiency analysis of multihypothesis motion-compensated prediction for video coding. **IEEE Transactions on Image Processing**, Institute of Electrical and Electronics Engineers (IEEE), v. 9, n. 2, p. 173–183, 2000. Disponível em: <https://doi.org/10.1109/83.821595>.
- GROIS, D. et al. Performance comparison of h.265/MPEG-HEVC, VP9, and h.264/MPEG-AVC encoders. In: **2013 Picture Coding Symposium (PCS)**. IEEE, 2013. Disponível em: <https://doi.org/10.1109/pcs.2013.6737766>.
- HAN, J. et al. A technical overview of AV1. **Proceedings of the IEEE**, Institute of Electrical and Electronics Engineers (IEEE), v. 109, n. 9, p. 1435–1462, set. 2021. Disponível em: <https://doi.org/10.1109/jproc.2021.3058584>.

- HEO, G. et al. An experimental study of a data compression technology-based intelligent data acquisition (IDAQ) system for structural health monitoring of a long-span bridge. **Applied Sciences**, MDPI AG, v. 8, n. 3, p. 361, mar. 2018. Disponível em: <https://doi.org/10.3390/app8030361>.
- LAUDE, T. et al. A comprehensive video codec comparison. **APSIPA Transactions on Signal and Information Processing**, Now Publishers, v. 8, n. 1, 2019. Disponível em: <https://doi.org/10.1017/atsip.2019.23>.
- MARTINEZ, B. et al. Sparse and random sampling techniques for high-resolution, full-field, BSS-based structural dynamics identification from video. **Sensors**, MDPI AG, v. 20, n. 12, p. 3526, jun. 2020. Disponível em: <https://doi.org/10.3390/s20123526>.
- MCNEILL, S.; ZIMMERMAN, D. A framework for blind modal identification using joint approximate diagonalization. **Mechanical Systems and Signal Processing**, Elsevier BV, v. 22, n. 7, p. 1526–1548, out. 2008. Disponível em: <https://doi.org/10.1016/j.ymssp.2008.01.010>.
- MEDEIROS, C. J. F. de; COSTA, J. A. F. Uma comparação empírica de métodos de redução de dimensionalidade aplicados a visualização de dados. **Learning Nonlinear Models**, SBRN, v. 6, n. 2, p. 81–110, 2008.
- PUNCHIHEWA, A.; BAILEY, D. A review of emerging video codecs: Challenges and opportunities. In: **2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)**. IEEE, 2020. Disponível em: <https://doi.org/10.1109/ivcnz51579.2020.9290536>.
- RASSOOL, R. VMAF reproducibility: Validating a perceptual practical video quality metric. In: **2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)**. IEEE, 2017. Disponível em: <https://doi.org/10.1109/bmsb.2017.7986143>.
- RIPPEL, O. et al. **Learned Video Compression**. arXiv, 2018. Disponível em: <https://arxiv.org/abs/1811.06981>.
- SILVA, I. G. O. **AValiação Experimental da Compressão de Vídeos de Estruturas Cíveis Utilizando Técnicas de Análise Modal Belém 2022**. [S.l.]: GitHub, 2022. <https://github.com/iurysilva/Research-Vibration-Video-Compression>.
- SILVA, M. et al. Spatio-temporal decomposition of 2d travelling waves from video measurements. **Mechanical Systems and Signal Processing**, Elsevier BV, v. 139, p. 106599, maio 2020. Disponível em: <https://doi.org/10.1016/j.ymssp.2019.106599>.
- SILVA, M.; GREEN, A.; MASCAREÑAS, D. A generalized technique for full-field blind identification of travelling waves and complex modes from video measurements with hilbert transform. In: **Data Science in Engineering, Volume 9**. Springer International Publishing, 2021. p. 233–236. Disponível em: https://doi.org/10.1007/978-3-030-76004-5_26.
- SONY, S.; LAVENTURE, S.; SADHU, A. A literature review of next-generation smart sensing technology in structural health monitoring. **Structural Control and Health Monitoring**, Wiley, v. 26, n. 3, p. e2321, jan. 2019. Disponível em: <https://doi.org/10.1002/stc.2321>.
- VANNE, J. et al. Comparative rate-distortion-complexity analysis of HEVC and AVC video codecs. **IEEE Transactions on Circuits and Systems for Video Technology**, Institute of Electrical and Electronics Engineers (IEEE), v. 22, n. 12, p. 1885–1898, dez. 2012. Disponível em: <https://doi.org/10.1109/tcsvt.2012.2223013>.

WADHWA, N. et al. Phase-based video motion processing. **ACM Transactions on Graphics**, Association for Computing Machinery (ACM), v. 32, n. 4, p. 1–10, jul. 2013. Disponível em: <https://doi.org/10.1145/2461912.2461966>.

WANG, Z.; SIMONCELLI, E.; BOVIK, A. Multiscale structural similarity for image quality assessment. In: **The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers**, 2003. IEEE, 2003. Disponível em: <https://doi.org/10.1109/acssc.2003.1292216>.

WU, C.-Y.; SINGHAL, N.; KRÄHENBÜHL, P. **Video Compression through Image Interpolation**. arXiv, 2018. Disponível em: <https://arxiv.org/abs/1804.06919>.

YANG, Y. et al. Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. **Mechanical Systems and Signal Processing**, Elsevier BV, v. 85, p. 567–590, fev. 2017. Disponível em: <https://doi.org/10.1016/j.ymssp.2016.08.041>.

YANG, Y.; NAGARAJAIAH, S. Blind modal identification of output-only structures in time-domain based on complexity pursuit. **Earthquake Engineering & Structural Dynamics**, Wiley, v. 42, n. 13, p. 1885–1905, maio 2013. Disponível em: <https://doi.org/10.1002/eqe.2302>.

ZHANG, T.; MAO, S. An overview of emerging video coding standards. **GetMobile: Mobile Computing and Communications**, Association for Computing Machinery (ACM), v. 22, n. 4, p. 13–20, maio 2019. Disponível em: <https://doi.org/10.1145/3325867.3325873>.