DISSERTAÇÃO DE MESTRADO

AVALIAÇÃO DA QUALIDADE DE VÍDEOS COM DEGRADAÇÕES TEMPORAIS

André Henrique Macedo da Costa

Brasília, dezembro de 2021

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO

AVALIAÇÃO DA QUALIDADE DE VÍDEOS COM DEGRADAÇÕES TEMPORAIS

André Henrique Macedo da Costa

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

Banca Examinadora

Profa. Mylène Christine Queiroz de Farias, ENE/UnB
Orientador _____

Prof. Daniel Guerreiro e Silva, ENE/UnB
Co-orientador _____

Daniel Costa Araújo, FTD/UnB
Examinador interno _____

Carlos Alexandre Barros de Mello, UFPE
Examinador externo _____

Dedicatória

Aos meus pais que são as pessoas que mais me deram suporte nessa jornada.

André Henrique Macedo da Costa

Agradecimentos

Existem muitas pessoas que eu gostaria de agradecer e tenho certeza que meus sentimentos não caberiam em apenas uma página. Gostaria de agradecer a minha família por todo o amor, atenção e carinho que recebi. Gostaria de agradecer aos meus avós maternos, Antônio e Teresa, por me mostrarem que é possível vencer as mais duras provações. Vocês foram um exemplo de trabalho duro e amor sem igual. Em especial, gostaria de agradecer aos meus pais, Pedro e Teresa, que provavelmente são as pessoas mais responsáveis pelas vitórias que tive na vida. Obrigado pela paciência e pelos ensinamentos constantes apesar dos meus esquecimentos. Vocês são responsáveis pelas melhores memórias que eu consigo me lembrar. Agradeço a minha irmã, Camila, pelas boas ferramentas de escrita. Elas foram muito utilizadas nesta tese.

Agradeço aos amigos que se mantiveram ao meu lado durante todos esses anos. A amizade de vocês foi sempre muito importante para mim. Vocês me ensinaram muitas coisas ao longo desses anos. Agradecimentos especiais para o João, a Paula e o Mikael. Obrigado pelas conversas, pelos lanches e as caminhadas no parque durante a quarentena. Nossos encontros foram uma terapia nesses tempos conturbados. Não poderia me esquecer dos amigos que fiz no GPDS. Vocês me salvaram diversas durante o mestrado e sempre me deram boas ideias. Gostaria de agradecer aos amigos que fiz no caratê. Nesses 17 anos de treinos, vocês foram responsáveis por manter minha cabeça no lugar certo.

Por último, mas não menos importante gostaria de agradecer a todos os professores que tive. Aprendi muito com os senhores. Dentre esses incríveis professores, gostaria de destacar os meus orientadores. Obrigado Mylène e Daniel por todos os ensinamentos durante esses últimos três anos. Obrigado pelo apoio neste semestre muito difícil. Foram muitas complicações, mas deu tudo certo. Obrigado por toda atenção que vocês me deram mesmo com todas as dificuldades da transição para o trabalho remoto. Ser orientando de vocês foi um enorme prazer.

André Henrique Macedo da Costa

RESUMO

Nesta dissertação, estudamos a avaliação de qualidade de vídeos contendo degradações temporais. De forma geral, a avaliação de qualidade de vídeos é realizada através de experimentos subjetivos ou métricas objetivas. Experimentos subjetivos ou psico-físicos são experimentos nos quais diferentes pessoas assistem a sequências de vídeos e atribuem notas de qualidade (ou outro atributo) para cada sequência. Apesar de experimentos subjetivos serem considerados mais precisos, eles possuem um alto custo, exigindo recursos físicos e tempo para serem executados. Métricas objetivas são algoritmos computacionais que estimam a qualidade de vídeos avaliando as suas propriedades físicas. O foco desta dissertação é o estudo e aprimoramento da métrica No-reference Autoencoder VidEo (NAVE). A NAVE é uma métrica de qualidade de vídeo sem referência (cega) que é baseada em *autoencoders* treinados para avaliações de degradações espaciais e temporais em vídeos. Estudos preliminares mostraram que os atributos temporais utilizados pela NAVE não estavam contribuindo para o desempenho da métrica. Desta forma, nesta dissertação fizemos um estudo de um conjunto de novos atributos espaciais e temporais que podem sr adicionados ao conjunto de atributos da NAVE, de forma a melhorar o seu desempenho. Realizamos diferentes testes e análises para verificar o desempenho da NAVE após a adição dos novos atributos. analisando o seu comportamento para degradações de compressão, perda de pacotes e congelamento de quadros. Os resultados apresentados apontam que a adição de atributos temporais mais sensíveis permite detectar melhor as degradações temporais e, consequentemente, melhorar o desempenho da métrica.

ABSTRACT

In this dissertation, we studied the quality assessment of videos containing temporal degradations. In general, video quality assessment is performed through subjective experiments or objective metrics. Subjective or psycho-physical experiments are experiments in which different people watch video sequences and assign quality scores (or other attributes) to each sequence. Although subjective experiments are considered more accurate, they are expensive, requiring physical resources and time to run. Objective metrics are computational algorithms that estimate the quality of videos by evaluating their physical properties. The focus of this dissertation is the study and improvement of the No-reference Autoencoder VidEo (NAVE) metric. NAVE is a No-Reference (blind) video quality metric that is based on *autoencoders* trained to evaluate spatial and temporal degradations in videos. Preliminary studies showed that the temporal attributes used by NAVE were not contributing to the performance of the metric. Thus, in this dissertation we studied a set of new spatial and temporal attributes that can be added to the NAVE's set of attributes, in order to improve its performance. We performed different tests and analyzes to verify the performance of the NAVE after adding the new attributes, analyzing its behavior for compression degradations, packet loss and frame freezing. Results show that the addition of more descriptive temporal attributes allows for a better detection of temporal degradations and, consequently, improves the performance of the metric.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS

Símbolos Gregos

QoS	Qualidade de Serviço
QoE	Qualidade de Experiência
NAVE	No-reference Autoencoder VidEo
FR	Métricas com Referência
RR	Métricas com Referência Reduzida
NR	Métricas sem Referência
RMSE	Erro Médio Quadrático
PSNR	Razão sinal-ruído de pico
NSS	Estatísticas de cenas naturais
AE1	Autoencoder 1
AE2	Autoencoder 2
TI	Temporal Index
SI	Spatial Index
MSCN	Mean Subtracted Contrast Normalized
GGD	Distribuição Generalizada Gaussiana
AGGD	Distribuição Assímetrica Generalizada Gaussiana
AVC	Advenced Video Coding
HEVC	High Efficency Video Coding
MOS	Mean Opinion Score
PCC	Coeficiente de Correlação de Pearson
SCC	Coeficiente de Correlação de Spearman

Capítulo 1

Introdução

1.1 Contextualização

Na última década, tem havido um crescimento no consumo de conteúdo de vídeo online. Grande parte deste crescimento tem acontecido devido à maior disponibilidade de plataformas de vídeo sob demanda como Netflix, HBO Max, Disney+ e outros. Recentemente, devido à pandemia da COVID-19, o fluxo de dados na internet sofreu um rápido crescimento devido ao grande número de pessoas trabalhando remotamente, o que levou ao crescimento do número de tele-conferências. Além disso, houve um aumento no consumo de conteúdo de vídeo online, o que aumentou a importância dos algoritmos para estimação da qualidade de vídeo.

Tradicionalmente, a medida de qualidade utilizada na avaliação de vídeos tem sido feito através de medidas de Qualidade de Serviço (QoS). As medidas de qualidade de serviço realizam a aferição da qualidade a partir de parâmetros que medem dados técnicos da rede de dados físicos como por exemplo, perdas de pacote e taxa de transmissão de bits. No entanto, nos últimos anos tem havido uma transição de métricas baseadas em qualidade de serviço para métricas ou abordagens baseadas em qualidade experiência (do inglês, Quality of Experience - QoE). A qualidade de experiência mede a qualidade levando em consideração aspectos da percepção humana. Mais especificamente, as medidas de qualidade de experiência se baseiam fortemente em trabalhos que levam em conta aspectos do sistema visual humano, assim como aspectos comportamentais humanos [3]. Desta forma, ao aferir a qualidade de um sinal de vídeo, é importante levar em consideração não apenas as taxas de compressão e transmissão, mas como as degradações nestes cenários são percebidas por seres humanos.

A avaliação de qualidade de vídeo pode ser feita de duas maneiras. Primeiramente, a avaliação pode ser feita através de experimentos psicofísicos ou experimentos subjetivos. Nesses tipos de experimentos, pessoas assistem a sequências de vídeo e atribuem julgamentos de qualidade para cada sequência assistida. Esse tipo de experimento tem que ser realizado em locais especiais, onde o ambiente é controlado, para que não haja interferências externas que comprometam os dados coletados. Devido ao método de coleta, os dados possuem uma alta fidedignidade com a percepção de qualidade humana. Em contrapartida, a obtenção dos dados nesse tipo de experimento é

extremamente custoso. A necessidade de que múltiplas pessoas assistam a múltiplas sequências de vídeo em um ambiente controlado é a principal desvantagem desse tipo de abordagem.

A segunda forma de avaliação de qualidade são os métodos objetivos de avaliação de qualidade, também chamados de métricas de qualidade, que consistem em algoritmos computacionais que tentam modelar a percepção de qualidade. A utilização de algoritmos computacionais na avaliação de qualidade permite que a avaliação seja feita de forma mais rápida e eficiente dispensando a necessidade de que diversas pessoas assistam a todas as sequências de teste. A implementação destes algoritmos, de forma geral, utiliza uma etapa de extração de atributos visuais e a uma função de mapeamento. Frequentemente, técnicas de aprendizado de máquinas são utilizadas para mapear os atributos extraídos em valores de qualidade de vídeo. Com base na disponibilidade do conteúdo original para comparação com o conteúdo, as métricas de qualidade podem ser divididas em três tipos. O primeiro tipo são as métricas com referência (no inglês *Full-Reference* - FR). Esse tipo de métricas utiliza integralmente o conteúdo original para realizar a comparação com o conteúdo avaliado. O segundo tipo de métrica são as métricas com referência reduzida (no inglês *Reduced Reference* - RR). Esse tipo de algoritmo não utiliza o conteúdo original na sua integralidade, mas apenas atributos representativos do vídeo que são comparados com os atributos correspondentes do vídeo original.

Finalmente, as métricas sem referência (no inglês *No-Reference* - NR) realizam a avaliação da qualidade do vídeo sem ter nenhum conhecimento sobre o conteúdo origem. O seu trabalho é identificar e extrair informações da sequência de vídeo avaliada que possam estatisticamente comprovar a presença ou não de degradação. Não possuir o conteúdo original à disposição torna a tarefa mais difícil, no entanto, estas são as condições mais comumente encontradas em cenários reais de transmissão. Isto torna as métricas sem referências mais adequadas do que as demais métricas para avaliação de qualidade em cenários de transmissão em tempo real.

Em um passado recente, a transmissão de vídeos via Internet não era feita em tempo real e as métricas de qualidade tinham que observar e levar em consideração apenas degradações espaciais. Degradações espaciais afetam apenas a correlação entre os pixels dentro de um mesmo quadro e não interfere entre quadros diferentes. Devido ao surgimento de transmissão de vídeos em tempo real, surgem também novos problemas e dentre esses problemas estão os defeitos temporais. Degradações temporais perturbam a informação de múltiplos quadros ou a correlação entre a informação de múltiplos quadros. Para a aferição da qualidade de vídeos com degradações temporais, é necessário desenvolver métricas de qualidade capazes de detectar tais degradações.

1.2 Definição do problema

O problema alvo deste trabalho é a melhoria do desempenho de uma métrica de qualidade sem referência para a avaliação de qualidade de vídeos com degradações temporais.

1.3 Objetivos do projeto

Esse trabalho tem como objetivo geral melhorar o desempenho da métrica sem referência de qualidade de vídeo *No-reference Autoencoder VidEo* (NAVE) [2] [4] no que diz respeito às degradações temporais. Seu trabalho é identificar a qualidade de um vídeo sem ter nenhum acesso ao conteúdo original do vídeo. Desta forma, a métrica deve extrair atributos dos quadros do vídeo que permitam distinguir entre sequências com degradações e sequências sem degradações. Além disso, a métrica tem que ser capaz de reconhecer gradações das degradações e diferenciar vídeos com quantidades diferentes de degradações.

O foco das melhorias da métrica diz respeito à avaliação de degradações temporais. Degradações temporais são degradações que não afetam necessariamente o conteúdo de cada quadro, mas perturbam a correlação entre diferentes quadros. Um exemplo bastante ilustrativo é o congelamento de quadros. O congelamento de quadros é uma degradação que ocorre quando a taxa média de transmissão é insuficiente para entregar os quadros do vídeo em tempo hábil para a sua reprodução. Neste cenário, a aplicação de vídeo reproduz novamente um quadro já recebido até que outro quadro chegue. Nesta situação, não existe nada de errado com o conteúdo do quadro, mas a cena é interrompida e o conteúdo perde correlação à medida que o tempo passa.

No intuito de atingir o objetivo geral, foram definidos os seguintes objetivos específicos:

- Selecionar novos atributos capazes de detectar degradações temporais;
- Analisar o desempenho dos novos atributos em comparação com os atributos já utilizados;
- Analisar as possíveis combinações de atributos e as interações entre eles;
- Avaliar a capacidade dos atributos de detectar degradações temporais;
- Avaliar o comportamento dos atributos temporais;
- Testar os modelos propostos e compará-los com os resultados de modelos já existentes.

1.4 Apresentação do manuscrito

No Capítulo 2, é feita a apresentação de conceitos básicos sobre o tema de estudo. São apresentados de forma resumida, conceitos sobre o aparelho visual humano, sistemas visual digitais e métricas de avaliação da qualidade de vídeos. Em seguida, o Capítulo 3 descreve a metodologia empregada no desenvolvimento do projeto. São apresentadas a métrica utilizada neste trabalho além dos atributos temporais e espaciais utilizados. Outro ponto apresentado são as bases de dados utilizadas no treinamento e teste da métrica proposta. Resultados experimentais são discutidos no Capítulo 4. No Capítulo 5, são feitos as conclusões do resultados obtidos e a proposta de trabalhos futuros.

Capítulo 2

Conceitos Básicos

Neste capítulo, são introduzidos conceitos básicos para o entendimento da dissertação. Primeiramente, é feita uma breve apresentação do sistema visual humano. Em seguida, é feita uma descrição sobre a estrutura básica de um sistema visual digital, incluindo conceitos de codificação de vídeo e degradações de vídeo. Por último, apresentamos os conceitos básicos sobre avaliação de qualidade de vídeo e algumas métricas de qualidade de vídeo populares.

2.1 Sistema Visual Humano

O sistema visual humano é responsável pelo sentido da visão que é um dos sentidos mais importantes na interação do ser humano com o mundo. Dentre as muitas tarefas executadas pelo sistema visual humano, estão a percepção de intensidade luminosa, a percepção de cores, a percepção de profundidade, a identificação de objetos, a identificação de movimentos e a identificação de padrões. De forma resumida, o sistema visual humano é composto pelo olho e pelo sistema nervoso central. O olho é uma estrutura aproximadamente esférica de cerca de 20 mm de diâmetro. A Figura 2.1 apresenta a corte transversal do olho humano e dos seus diversos tecidos.

Durante o processo da visão, a luz passa primeiramente pela córnea, que é a membrana exterior do olho. Ao passar pela córnea, a luz entra na câmara anterior composta por humor aquoso e adentra passando pela pupila. A pupila é um orifício no centro do olho cuja abertura é controlada pela íris e por onde a luz passa. A íris é uma estrutura que se expande e se contrai de forma a controlar a quantidade de luz que entra na câmara interior do olho. O funcionamento da íris é semelhante ao diagrama das câmeras fotográficas. Após passar pela pupila, a luz deve passar pela lente do olho chamada cristalino. Esta lente controla a direção que a luz atinge a região interna do olho, a retina. Antes da retina, existe a câmara interior do olho que é composta pelo humor vítreo.A superfície interna do olho é chamada de retina. Ela é composta por diferentes tipos de células fotossensíveis chamadas de cones e bastonetes. Os bastonetes são células fotossensíveis presentes na retina dos seres humanos que possuem a principal função da detecção de intensidade luminosa sendo incapazes de identificar cores. Eles também são utilizados na visão noturna e na visão periférica. No olho humano, existem cerca de 120 milhões de bastonetes e os bastonetes são

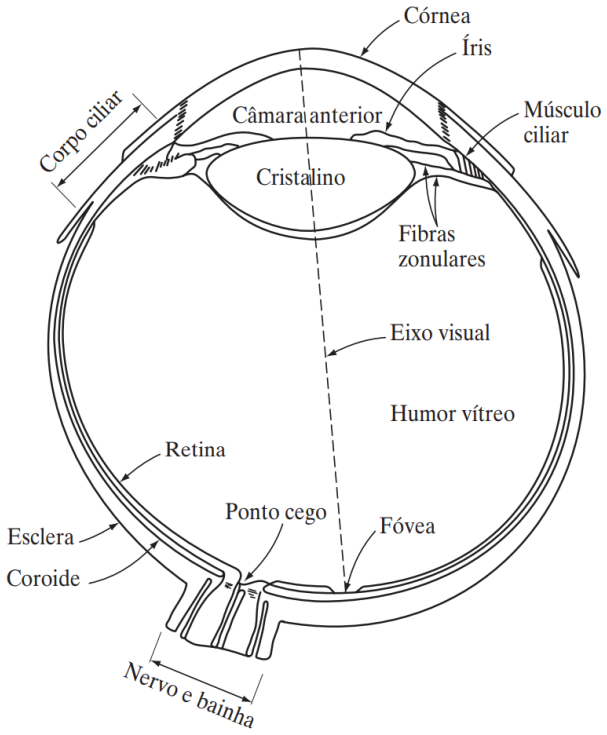


Figura 2.1: Diagrama simplificado do corte transversal do olho humano [1].

aproximadamente 100 vezes mais sensíveis à luz do que os cones. Os cones são células fotossensíveis capazes de identificar as diferentes cores do espectro da luz visível. Nos seres humanos existem três tipos de cone. O primeiro tipo possui a capacidade de identificar cores vermelhas, ou seja, ondas eletromagnéticas com comprimento de onda de pico de aproximadamente 564 e 580nm. O segundo tipo de cones é responsável por visualizar as cores de tons verdes com comprimento de onda de pico entre 534 e 545 nm, enquanto o terceiro tipo de cone é responsável por visualizar as cores azuladas que possuem comprimento de onda de pico entre 420 e 440 nm.

2.2 Sistemas Visuais Digitais

De forma resumida, os sistemas visuais digitais podem ser divididos em três partes: transmissor, canal e receptor. O transmissor é responsável por capturar, codificar e enviar o sinal. O receptor é responsável por receber, decodificar e reproduzir o sinal. O canal é o meio pelo qual o sinal é transmitido. Este trabalho tem como objeto de estudo os sinais de vídeo. A seguir serão apresentados conceitos básicos de codificação de vídeo e degradações de vídeo.

2.2.1 Compressão de Vídeos

Algoritmos de compressão são de extrema importância para o bom funcionamento da Internet, especialmente quando se fala da transmissão de vídeo. Estes algoritmos são responsáveis por eliminar redundâncias no conteúdo e reduzir o espaço de armazenamento necessário dos dados. Os

algoritmos de compressão podem ser divididos em compressão com perda, em inglês *lossy compression*, ou compressão sem perdas, em inglês *losless compression*. Algoritmos de compressão sem perdas possuem maior resistência ao corrompimento dos dados, porém possuem uma baixa taxa de compressão. Algoritmos de compressão com perdas oferece taxas de compressão maiores, porém abrem a possibilidade de que haja degradações no conteúdo. As degradações provenientes da compressão podem ser visíveis ou não visíveis. A maioria dos algoritmos utilizados na compressão de vídeos são algoritmos com perdas, como por exemplo, os algoritmos H.264 [5] e H.265 [6].

Apesar da qualidade dos codificadores modernos, falhas na transmissão dos dados ou uma compressão muito severa podem introduzir diversos tipos de artefatos ou degradações ao conteúdo do vídeo. Esse problema se torna ainda mais complexo se o vídeo está sendo transmitido ao vivo, o que impede a possibilidade de reenvio da informação. Assim, métricas de qualidade podem ser utilizadas para medir a qualidade do vídeo transmitido e para ajustar a compressão de acordo. Dentre as possíveis degradações que podem afetar o sinal de vídeo, podemos elencar as seguintes:

- Blocagem
  - Os artefatos de blocagem consistem em descontinuidades em formato de blocos no quadro causados pela compressão. Este efeito acontece devido à compressão ser realizada separadamente em macro blocos (conjunto de pixels) ao invés de ser realizada no quadro inteiro. O efeito visual deste artefato é aparição dos macro blocos utilizados na compressão [7]. A Figura 2.2 apresenta um exemplo de quadro original e o mesmo quadro contendo a degradação de blocagem.

Original Blocagem

Figura 2.2: Exemplo de par de quadros de um vídeo: (esquerda) versão original e (direita) versão com a degradação de blocagem.

- Borrado
  - A degradação de borrado consiste na redução dos níveis de detalhes espaciais do quadro. Ela está entre as degradações mais comuns, estando relacionada à perda ou redução das informações de alta frequência [8]. A Figura 2.3 apresenta um exemplo de quadro original e o mesmo quadro contendo a degradação de borrado.

Original Borrado

Figura 2.3: Exemplo de par de quadros de um vídeo: (esquerda) versão original e (direita) versão com a degradação de borrado.

- Congelamento de Quadros
  - O cenário de congelamento de quadros ocorre quando a taxa de transmissão da rede é inferior a taxa de reprodução do conteúdo. Devido ao desequilíbrio entre as duas taxas, o sistema de reprodução tem que aguardar a chegada de mais conteúdo para que seja possível retomar a reprodução. Durante esse intervalo de espera, o sistema é forçado a repetir a reprodução do último conteúdo recebido. Esta degradação se torna mais danoso se o áudio da cena continuar a ser reproduzido enquanto a cena se encontra estática.
- Perda de Pacotes
  - A perda de pacotes acontece quando a taxa de transmissão do sinal supera a capacidade de reprodução do conteúdo. Sendo assim, o *buffer* da aplicação de vídeo no receptor é preenchido completamente e o sistema é obrigado a descartar os pacotes que estão sendo recebidos em seguida. Devido a esse descarte ocorre a eliminação de parte do conteúdo futuro. A Figura 2.4 apresenta um exemplo de quadro original e o mesmo quadro contendo a degradação de perda de pacotes.

Original Perda de pacotes

Figura 2.4: Exemplo de par de quadros de um vídeo: (esquerda) versão original e (direita) versão com a degradação de perda de pacotes.

2.3 Avaliação da Qualidade de Vídeos

A qualidade de vídeo pode ser obtida de duas formas: experimentos subjetivos ou métricas objetivas de qualidade. Os experimentos subjetivos são experimentos onde pessoas avaliam diferentes sequências de vídeo e atribuem a cada uma delas um valor de qualidade. Esse tipo de experimento é capaz de obter um valor muito fidedigno da percepção humana de qualidade visual. O ônus desse tipo de prática são o tempo e o custo de realizar os experimentos, devido à necessidade de que diversas pessoas assistam a conjuntos de vídeos por um período considerável de tempo. A segunda opção é uma métrica objetiva de qualidade. Métricas de qualidade são algoritmos que tentam modelar matematicamente a percepção de qualidade visual. Elas não apresentam os custos elevados de tempo e pessoas, no entanto, é fácil perceber que tal tarefa é complexa e por isso diversos trabalhos científicos vêm sendo realizados nesse campo.

Como já dito, obter a análise de qualidade de vídeo via experimentos subjetivos não é uma abordagem prática para ser amplamente utilizada, porém essa abordagem possui uma importância muito grande na área de qualidade de vídeo. Experimentos subjetivos são a melhor forma de se capturar a percepção humana de qualidade. Essa captura é importante para realizar a criação de bancos de dados para a criação de métricas de qualidade e para realizar a modelagem da percepção humana. A modelagem da percepção humana é muito importante pois traz a possibilidade de que algoritmos dos sistemas digitais de vídeo sejam melhorados de forma a melhora atender as necessidades humanas com uma menor quantidade de recursos. Lembrando que a análise e modelagem da percepção humana é um tema complexo e multi-disciplinar. E, embora a compreensão do sistema visual humano ainda seja um desafio, várias características do sistema já foram descobertas, como por exemplo o efeito de histeresis **??** e a recência [9].

2.3.1 Avaliação Objetiva de qualidade de vídeo

Métricas objetivas de qualidade podem ser utilizadas para medira qualidade do vídeo que compõe a qualidade de experiência, *quality of experience* (QoE). Na sua maioria, as métricas de qualidade são compostas de duas partes: extração de descritores e modelo de decisão. Quanto a forma da extração dos descritores, conforme mencionado no capítulo anterior, as métricas podem ser divididas em 3 tipos:

- Métricas com Referência ou *Full-Reference*, FR;
- Métricas com Referência Reduzida ou *Reduced-Reference*, RR;
- Métricas sem Referência ou *No-Reference*, NR.

A seguir detalhamos cada um destes tipos de métricas.

2.3.2 Métricas com Referência

As métricas FR utilizam duas versões do vídeo para realizar a análise de qualidade, uma versão original e sem degradação e uma versão com degradação. A métrica realiza a comparação entre os dois conteúdos para medir a qualidade da versão degradada. Dessa forma, é possível notar que a existência das duas formas de conteúdo é o principal fator limitante da implementação desse tipo de métrica e tal fator limitante dificulta a utilização desse tipo de métrica em tempo real. No geral, o desempenho desse tipo de métrica é superior aos demais tipos já que o conteúdo original é utilizado.

Dentre as métricas com referência mais populares, podemos citar o erro quadrático médio (em inglês *mean-squared error* - MSE) e a razão sinal-ruído de pico (em inglês *Peak Signal-to-Noise Ratio* - PSNR). Essas duas métricas são simples e realizam apenas a comparação dos dados dos quadros de cada vídeo sem levar em consideração o conteúdo da cena. Por esse motivo, elas não possuem a melhor correlação com a percepção de qualidade humana [10] [11], no entanto, elas são rápidas e simples de serem implementadas.

O erro quadrático médio é calculado da seguinte forma:

$$MSE = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(I_o(m,n) - I_d(m,n))^2}{M \cdot N}, \tag{2.1}$$

onde $I_o(m,n)$ é a imagem original e $I_d(m,n)$ a imagem degradada. A razão sinal-ruído de pico é calculada através da seguinte fórmula:

$$PSNR = 10 \cdot \log_{10} \frac{\text{Max}(I(x,y))}{\sqrt{MSE}}. \tag{2.2}$$

Outra métrica de qualidade com referência bastante popular é o Índice de Similaridade Estrutural (*structural similarity index measure* - SSIM) [12]. Originalmente desenvolvida como uma métrica de qualidade de imagens, a SSIM pode ser aplicada a cada quadro de um vídeo e operar como métrica de qualidade de vídeo. A SSIM compara dois quadros x e y através da equação:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_x^2 + C_2)}, \tag{2.3}$$

onde x é o quadro do vídeo original, μ_x é a média do quadro x, σ_x é a variância do quadro x, y é o quadro do vídeo degradado, μ_y é a média do quadro y, σ_y é a variância do quadro y e C_1 e C_2 são constantes para evitar uma divisão por zero.

2.3.3 Métricas com Referência Reduzida

Métricas com Referência Reduzida fazem a comparação parcial entre o conteúdo original e o conteúdo degradado. Diferentemente das métricas FR, as métricas RR realizam a extração de atributos dos conteúdos degradado e não-degradado e realizam a avaliação de qualidade a partir da comparação dos atributos extraídos. Dentre essa classe de métricas, podemos citar a PeQASO [13].

2.3.4 Métricas Sem Referência

Métricas Sem-Referência (NR) diferem das métricas anteriores quanto à disponibilidade de conteúdo original. Ao contrário dos outros tipos de métricas, métricas NR não utilizam informações sobre o sinal original e, portanto, elas são as únicas que podem ser utilizadas em um cenário de avaliação em tempo real. Devido à provável impossibilidade do envio de uma versão sem degradação do sinal original, métricas NR são responsáveis por fazer a medida de qualidade na maior parte das transmissões de vídeo. Logo, fica evidente a importância do desenvolvimento e melhoramento destas métricas.

Uma vez que a comparação entre o sinal original e o sinal degradado é impossível, as métricas NR precisam detectar a presença ou não de degradações analisando apenas o sinal recebido ou de teste. Essa detecção pode ser feita de diversas formas, dentre elas existe um grupo de métricas que utiliza estatísticas de cenas naturais (em inglês *Natural Scene Statistics* - NSS)[14]. Dentre as métricas sem referência que utilizam a abordagem NSS podemos citar as seguintes: PTQM [15], VIIDEO [16] e NAVE [2]. Essas métricas consideram que diferentes degradações alteram de forma diferente a distribuição dos coeficientes normalizados de luminância [17]. Desta forma, degradações alteram o formato da distribuição gaussiana dos coeficientes, com cada tipo de degradação tendo uma curva de distribuição específica.

2.4 Métricas de Qualidade de Vídeo com Aspectos Temporais

Dentro avaliação de qualidade de vídeo, existe uma assimetria grande entre a identificação de degradações espaciais e temporais. Degradações espaciais são caracterizadas por uma redução da correlação entre os pixels de um quadro enquanto degradações temporais são caracterizadas por uma redução da correlação temporal entre quadros do vídeo. Devido às características das degradações espaciais, as métricas de qualidade de vídeo fazem uso de muitas das técnicas utilizadas para a avaliação da qualidade de imagens. Ou seja, de certa maneira, as métricas de qualidade de imagens formam um arcabouço de técnicas que podem ser utilizadas para a avaliação de degradações espaciais em vídeos. Degradações temporais, ao contrário, são degradações nativas de vídeos e a priori não são capturadas pelo métricas de qualidade de imagens.

Recentemente, houve um aumento no interesse por degradações temporais devido à sua presença em sistemas de *streaming* de vídeo. Ainda se trata de um campo pouco explorado, mas já existem alguns trabalhos desenvolvidos que podem ser vistos a seguir. No trabalho *Video is a cube* [18], Keimel *et al* propõem uma abordagem da avaliação de qualidade de vídeo baseada em técnicas de análise de dados na qual se considera a multidimensionalidade da estrutura do vídeo. O vídeo é tratado como um cubo, uma estrutura tridimensional, onde é analisado como uma estrutura que pode ser segmentada. Atributos podem ser extraídos dessas subestruturas de forma a obter uma melhor captura da informação. As análises feitas neste trabalho apontam para um melhor desempenho na avaliação da qualidade de experiência quando se é considerado a dimensão do tempo durante a extração e processamento dos descritores. Outra análise feita diz respeito ao uso de *Principal Component Analysis* [19] em combinação com a regressão linear,

PCR, na sub-amostragem da informação extraída do vídeo. Essa abordagem obtém uma melhora de desempenho da métrica com um baixo custo computacional.

No trabalhos *Spatiotemporal of Naturalness* [20], os autores Sinno e Bovik propõem um conjunto de atributos de qualidade para degradações temporais. Os atributos são baseados na abordagem *Natural Scene Statitics*. Os atributos tentam fazer uso dos conceitos e conhecimentos de *Natural Scene Statistics* que detectam a presença de degradações espaciais a partir da diferença entre a distribuição gaussiana dos coeficientes calculados com os pixel de uma imagem. Os atributos propostos nesse trabalho partem do pressuposto de que essa modelagem da detecção das degradações espaciais pode ser aplicada na diferença entre quadros subsequentes para a detecção de degradações temporais.

A PeQASO [13] desenvolvida por Aabed *et al* é uma métrica de vídeo com referência reduzida que estima a qualidade de um vídeo contendo distorções espaço-temporais. A detecção das distorções utiliza o método Horn-Schunck de *optical flow* [21] em conjunto com a redução da dimensão dos quadros por um fator de 2 para calcular o *optical flow* em múltiplas escalas. Para cada quadro, cria-se um mapa de *optical flow* composto pelos valores obtidos no *optical flow* em 3 escalas espaciais diferentes. A qualidade do vídeo é obtida através da comparação entre o mapa de *optical flow* do vídeo degradado e o mapa de *optical flow* do vídeo original.

No trabalho de Jari Korhonen [22], intitulado *Two Level Approach for No-Reference Consumer Video Quality Assessment*, ele propõe uma métrica sem referência que utiliza um mecanismo de duas etapas para encontrar quadros de interesse para o cálculo dos atributos. O primeiro conjunto de atributos calculados é principalmente relacionado ao movimento da cena e são responsáveis pela captura da informação temporal do vídeo e por consequência, a informação sobre as degradações temporais. Além de serem utilizados na detecção de degradações, os atributos de baixa complexidade são utilizados para selecionar os quadros mais representativos. Um conjunto de atributos de alta complexidade são calculados nesse conjunto de quadros que é responsável pela detecção de degradações espaciais. Simultaneamente ao cálculo dos atributos de alta complexidade, é realizado um agrupamento de desvio padrão temporal, *temporal standard deviation pooling*, nos atributos de baixa complexidade para se calcular a atributos de consistência de movimento. Os conjuntos de atributos passam por um agrupamento de média temporal e são concatenados para formar o vetor de atributos. A estimação da nota de qualidade é feito por meio de um algoritmo SVR (*support machine regressor*).

Finalmente, os autores Vu e Chandler [23] desenvolveram a métrica com referência, denominada ViS3, que estima a qualidade de um vídeo através da detecção de distorções espaciais e espaço-temporais. O algoritmo é composto de duas etapas: detecção espacial e detecção espaço-temporal. A detecção espacial utiliza uma combinação do algoritmo MAD [24] para detectar a distorções em cada quadro e o método de *optical flow* Lucas–Kanade [25] que estima a quantidade de distorção espacial, ViS1. A etapa espaço-temporal calcula a dissimilaridade entre quadros espaço-temporais da componente de luminância, que são quadros formados por linhas ou colunas de diferentes quadros do vídeo. Os valores da dissimilaridade dos quadros espaço-temporais são combinados em um único valor que representa a qualidade espaço temporal do vídeo degradado, ViS2. Os dois

valores de qualidade são combinados para produzir uma nota de qualidade, ViS3, para o vídeo inteiro.

Capítulo 3

Metodologia

Neste capítulo, são apresentados os métodos utilizados neste trabalho. Tendo em vista que o objetivo desta dissertação é o melhoramento da métrica de qualidade NAVE, especificamente no que diz respeito a sua sensibilidade a degradações temporais, neste capítulo descrevemos os testes realizados com esta métrica, que consistem basicamente na adição de novos atributos espaciais e temporais ao conjunto de atributos já utilizados pela métrica. Os testes dos novos conjuntos de atributos foram realizado utilizando diferentes bases de dados, de forma a termos uma análise mais detalhada do comportamento da métrica modificada com a adição destes novos atributos.

3.1 NAVE

Métricas de qualidade sem referência possuem a difícil tarefa de aferir a qualidade de um vídeo em um tempo adequado e sem ter o conteúdo original à disposição. Por esse motivo, é necessário que o modelo seja rápido e possua capacidade de detectar as degradações mais relevantes para as aplicações em questão. O modelo deve ser capaz de extrair atributos sensíveis às degradações presentes no sinal de vídeo e, em seguida, atribuir uma nota de qualidade geral ao conteúdo. A métrica de qualidade de vídeo utilizada nesse trabalho é a *No-reference Autoencoder VidEo* (NAVE) [2]. A estrutura da NAVE é composta de três partes: extração de atributos, redução do número de atributos, estimação de qualidade.

O conjunto de atributos original da NAVE é composta pelos atributos da métricas DIIVINE [26] e os atributos TI, *Temporal Index*, e SI, *Spatial Index* [27]. A Redução do número de atributos é realizado através de 2 autoencoders para que se obtenha uma representação compacta dos atributos de qualidade. A função de classificação é uma função *softmax*. A NAVE foi desenvolvida para identificar a presença e medir a qualidade de vídeo com degradações de compressão H.264, H.265, perda de pacotes e congelamento de quadros. A arquitetura da NAVE é apresentada na Figura 3.1.

O primeiro autoencoder, denominado AE1, realiza a redução dos atributos para um vetor de tamanho $50 \times n$, onde n é o número de quadros do vídeo. O segundo autoencoder, AE2, recebe o vetor de saída do primeiro autoencoder e reduz o vetor $50 \times n$ para um vetor $20 \times n$. O treinamento

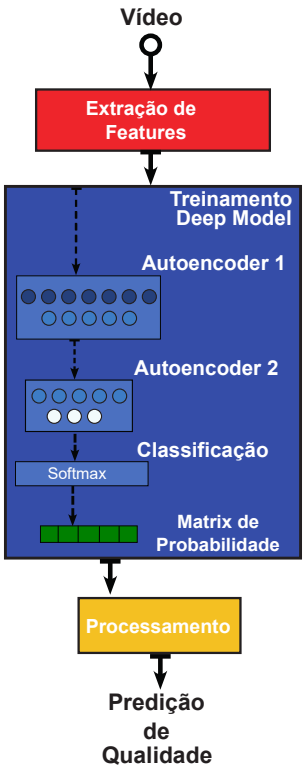


Figura 3.1: Diagrama da arquitetura da métrica *No-reference Autoencoder VidEo* (NAVE) [2].

de cada autoencoder é feito separadamente. Em um primeiro momento, o AE1 é treinado e, após o fim do seu treinamento, a atualização dos seus parâmetros é desabilitada para o treinamento do segundo autoencoder. Durante o treinamento do AE2, o AE1 realiza apenas a redução do vetor original e os parâmetros de AE2 são atualizados. Em ambos os treinamentos, são utilizados os seguintes parâmetros: o peso L2 de regularização é $0,001$, o coeficiente de regulação de esparsidade é igual a 4, o coeficiente de proporção de esparsidade é igual a $0,05$, uma função linear é utilizada no decodificador e 100 épocas de treinamento.

Após o treinamento dos autoencoders, o treinamento do classificador é realizado. O classificador mapeia o vetor de saída do AE2 para as notas de qualidade utilizando uma camada com função *Softmax*. Após o treinamento dessas 3 partes, os *encoders* dos 2 autoencoders e a camada de classificação são concatenadas para formar a estrutura completa da NAVE, denominada DeepNet. A DeepNet é treinada recebendo o vetor contendo todos os atributos e produzindo uma nota de qualidade para cada vídeo. Mesmo cada parte da DeepNet sendo treinada individualmente, o treinamento final é realizado de forma a refinar os parâmetros dos autoencoders e do classificador para operarem em conjunto. A Figura 3.2 mostra de forma resumida o treinamento da métrica NAVE.

3.2 Atributos

Neste trabalho, dentre os atributos originais da métrica NAVE, excluímos os atributos TI e SI e mantemos os atributos espaciais da métrica DIIVINE. O motivo da exclusão dos atributos TI e

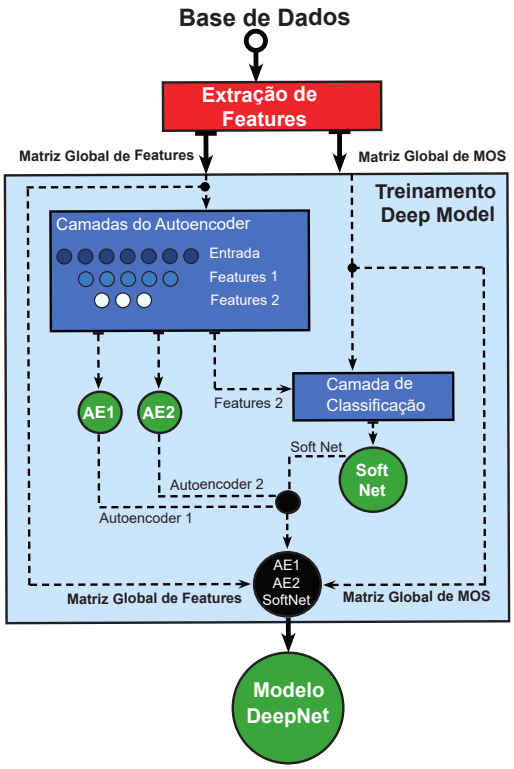


Figura 3.2: Diagrama contendo de forma resumida o treinamento da métrica NAVE [2].

SI é explicado na seção 4.1. Os atributos DIIVINE pertencem ao grupo de atributos baseados em NSS e foram originalmente desenvolvidos para estimação da qualidade de imagens. No entanto, uma prática comum na área de qualidade de vídeo consiste em utilizar atributos e métricas de imagens em cada um dos quadros do vídeo, combinando-as ou processando-as para o conjunto total de quadros. Essa estratégia é muito eficiente na avaliação de degradações espaciais, uma vez que este tipo de degradação afeta os pixels de cada quadro isoladamente. Por outro lado, as degradações temporais podem alterar a correlação entre os pixels de diferentes quadros. Existe um desequilíbrio entre o número de atributos espaciais e temporais na NAVE, o que pode levar a um pior desempenho na detecção de qualidade no que degradações de natureza temporal. A seguir serão descritas os atributos espaciais e temporais utilizadas neste trabalho.

3.2.1 Atributos DIIVINE

A DIIVINE [28] é uma métrica de qualidade de imagens utilizada na detecção de distorções espaciais. Originalmente utilizada na avaliação de qualidade de degradações, como borrado gaussiano, *fast-fadding*, ruído gaussiano e compressões JPEG e JPEG2000. Na NAVE, utilizamos os atributos desta métrica com o objetivo de detectar as degradações de compressão H.264 e H.265.

A métrica DIIVINE possui 88 atributos, formando um vetor de dimensão $88 \times n$, onde n é o número de quadros do vídeo. Esse conjunto de atributos é obtido através da transformada *wavelet* complexa piramidal manobrável, *Steerable Pyramid Complex Wavelet Transform*, que produz coeficientes complexos, Z_s^O em 6 orientações, $O \in \{0^o, 30^o, 60^o, 90^o, 120^o, 150^o\}$, diferentes para cada

uma das três escalas, $s \in \{1, 2, 3\}$. Todavia, são utilizadas apenas 2 escalas, $s \in \{1, 2\}$, já que a maioria dos tipos de distorção afetam mais a informação presente nas altas frequências e, dessa forma, a escala menor possui um conteúdo menos degradado [29]. Os coeficientes passam por uma normalização de forma a diminuir a dependência entre as sub-bandas [30]. A partir desses coeficientes normalizados, são obtidos vários subconjuntos de atributos.

O subconjunto de estatísticas de escala e orientação é composto pelos atributos f_1 a f_{24}. Estes atributos são obtidos através do ajuste da distribuição dos coeficientes de cada combinação de orientações e uma sub-banda a uma distribuição Gaussiana generalizada, conforme a seguinte equação:

$$f(x; \sigma^2, \gamma) = a \cdot e^{[-(b|x|)^\gamma]}, \tag{3.1}$$

onde γ é o parâmetro que controla o formato da distribuição e σ^2 é a variância da distribuição. Os parâmetros a, b são dados pelas seguintes equações:

$$a = \frac{b\gamma}{2\Gamma(1/\gamma)} \tag{3.2}$$

e

$$b = \sigma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}, \tag{3.3}$$

onde e a função Γ é descrita pela seguinte equação:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0. \tag{3.4}$$

O melhor ajuste de cada sub-banda produz 12 pares de coeficientes, $\gamma_{s,O}$ e $\sigma^2_{s,O}$. Em vídeos sem degradação esses 2 parâmetros produzem uma distribuição Gaussiana, enquanto que a presença de degradações produz distribuições com outros formatos.

Para obter os atributos de estatísticas de orientação f_{25} a f_{31}, uma concatenação dos coeficientes de sub-bandas de diferentes escalas é realizada, seguida pelo ajuste a uma Distribuição Gaussiana Generalizada, Generalized Gaussian Distribuition (GGD). Esse processo é necessário uma vez que as imagens tem estruturas em múltiplas escalas e, por consequência, as degradações também estão presentes nas diferentes escalas da imagem. Os atributos f_{25} a f_{30} são os coeficientes γ para cada uma das orientações, enquanto que o atributos f_{31} é o valor de γ para a concatenação dos coeficientes de todas as sub-bandas. O subconjunto de atributos f_{32} a f_{43} é composto pelos atributos de correlação entre escalas. Tais atributos são obtidos através do cálculo da correlação estrutural janelada [12] entre a resposta passa-alta e a resposta passa-baixa, que é expressa pela seguinte equação:

$$\rho = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{3.5}$$

onde σ_{xy} é a correlação entre as respostas do passa-alta e passa-baixa dentro da máscara do filtro, σ_x e σ_Y são as variâncias dentro de cada máscara e C_2 é uma constante para evitar divisões por zero. O valor de correlação é calculado para cada umas das 12 sub-bandas que produz os atributos f_{32} a f_{43}.

O subconjunto de atributos f_{44} a f_{73} é obtido através do cálculo da correlação espacial dentro da mesma sub-banda. Sabe-se que uma imagem sem degradações possui uma alta correlação

espacial (inter-pixels), ao contrário de uma imagem com degradações. Essa correlação pode ser obtida através do cálculo da diferença entre coeficientes separados por uma distância τ que é, posteriormente, ajustada a uma função polinomial de 3ª ordem. O ajuste é feito para cada uma das orientações para a escala $s = 1$, onde os coeficientes do ajuste e o erro entre o ajuste e a correlação são os atributos. O último subconjunto de atributos reúne os atributos obtidos a partir da informação presente nas estatísticas entre orientações adjacentes. Para cada combinação de orientações, a correlação estrutural entre as imagens das escalas $s = 1$ e $s = 2$ é calculada, conforme a Eq. 3.5. Dessa forma, são produzidas os atributos f_{74} a f_{88}.

3.2.2 Atributos de Sinno *et al.*

Neste trabalho, foram utilizados os atributos originais da métrica NAVE em conjunto com novos atributos propostas na literatura. Os primeiros atributos testados foram os atributos propostos por Sinno *et al.* [20], que são referenciados como atributos Sinno. Esses atributos fazem parte de uma família de atributos baseados na estatística de cenas naturais (do inglês *Natural Scene Statistics* - NSS). Essa família de atributos segue o princípio de que diferentes degradações alteram a forma da distribuição Gaussiana dos coeficientes MSCN (*Mean Subtracted Contrast Normalized Coefficients*). Esse conjunto de atributos foi escolhido devido a sua correlação com diversos tipos de degradações, incluindo degradações temporais, presentes na base de dados construída por Sinno *et al.* [31].

Os atributos Sinno são calculados subtraindo dois pixels de quadros subsequentes, sendo t o índice temporal e o par (i, j) os índices temporais correspondentes às informações verticais e horizontais. O primeiro passo, consiste em subtrair as intensidades dos pixels de quadros subsequentes nas 4 direções, obtendo os seguintes sinais diferenças:

$$D_H(i,j,t) = I_t(i,j) - I_{t+1}(i,j-1), \tag{3.6}$$

$$D_V(i,j,t) = I_t(i,j) - I_{t+1}(i-1,j), \tag{3.7}$$

$$D_{D1}(i,j,t) = I_t(i,j) - I_{t+1}(i-1,j-1), \tag{3.8}$$

$$D_{D2}(i,j,t) = I_t(i,j) - I_{t+1}(i-1,j+1). \tag{3.9}$$

Em seguida, são calculadas as médias locais e variâncias de cada quadro dos sinais diferenças. As duas medidas são calculadas utilizando uma máscara Gaussiana normalizada,

$$D^M_{(.)}(i,j,t) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} g_{m,n} D_{(.)}(i+m, j+n), \tag{3.10}$$

onde $(.) \in \{D_1, D_2, H, V\}$, $g_{m,n}$ é a máscara Gaussiana e M e N são respectivamente a altura e largura da máscara, que neste caso possuem ambas um valor igual a 3. A variância é calculada através da seguinte fórmula:

$$D^V_{(.)}(i,j,t) = \sum_{m=1}^{w} \sum_{n=1}^{h} (D_{(.)}(i,j,t) - D^M_{(.)}(i,j,t))^2, \tag{3.11}$$

onde $(.) \in \{D_1, D_2, H, V\}$ e w e h são as dimensões dos quadros. Dessa forma, temos a média calculada de forma local e a variância de forma global para cada quadro. Os valores acima são utilizados para calcular os coeficientes MSCN, onde se subtrai a média local da diferença entre quadros e se divide essa diferença pela variância do quadro:

$$D_{MSCN}(i,j,t) = \frac{D_{(.)}(i,j,t) - D^M_{(.)}(i,j,t)}{D^V_{(.)}(i,j,t) + 1}. \tag{3.12}$$

Após o cálculo dos coeficientes MSCN, os quadros são divididos em sub-blocos (*patches*) de tamanho 96×96, de forma semelhante ao que é feito na métrica NIQE [32]. Caso o tamanho do quadro não seja múltiplo das dimensões do *patch*, o processo descrito acima é realizado eliminando linhas ou colunas nas bordas do quadro, uma vez que o conteúdo mais importante do quadro encontra-se na parte central.

Os coeficientes MSCN de cada *patch* são ajustados a uma Distribuição Gaussiana Generalizada (GGD) utilizando o algoritmo proposto por Lasmar *et al.* [33]. O algoritmo realiza o ajuste utilizando a seguinte função de correspondência de momentos apresentada na equação 3.1. Um par de parâmetros γ e σ^2 que cria o melhor ajuste da distribuição é calculado para cada *patch* de cada quadro do vídeo.

Duas formas de pós-processamento foram utilizadas no conjunto de pares γ e σ^2 para produzir dois tipos de atributos Sinno. Escolheu-se por testar se haveria mudança caso fossem utilizados um vetor com todos valores γ e σ^2 de cada *patch* de cada quadro ou se fosse feita a média dentro de cada quadro (instante t), resultando em apenas um par de valores para cada quadro. A primeira forma de pós processamento produz o conjunto de atributos SinnoP, onde são concatenados os pares γ e σ^2 de todos os *patches* do quadro para formar o vetor de atributos referente aquele quadro. A segunda forma de pós processamento, SinnoM, utiliza apenas um par de parâmetros γ e σ^2 por quadros. O par de parâmetros do conjunto SinnoM é a média dos parâmetros de cada *patch* de cada quadro. Esse pós processamento foi realizado com o objetivo de verificar se haveria ganho de desempenho com o uso de um número maior de atributos ou se poderíamos ganhar em eficiência utilizando apenas 2 valores por quadro.

3.2.3 Atributos BRISQUE

O segundo conjunto de atributos utilizado foi o conjunto de atributos da métrica BRISQUE [34], que são atributos sem-referência baseados na abordagem NSS, com formulação matemática similar aos atributos Sinno. Primeiramente, os coeficientes MSCN dos pixels de cada quadro do vídeo são calculados, conforme a seguinte equação:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1}, \tag{3.13}$$

onde $\mu(i,j)$ é a média local centrada no pixel (i,j) e $\sigma(i,j)$ é a variância local centrada no pixel (i,j). A média local é calculada utilizando a seguinte equação:

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} g_{k,l} I(i+k, j+l), \tag{3.14}$$

na qual, $g_{k,l}$ é um filtro passa baixa Gaussiano de dimensões $(2K+1, 2L+1)$. Neste caso, $K = L = 3$.

Como mencionado anteriormente, os coeficientes MSCN do BRISQUE se diferem dos coeficientes MSCN dos atributos do trabalho de Sinno *et al.* no cálculo da variância. Ao contrário dos atributos Sinno, que utilizam a variância global do quadro, os atributos BRISQUE utilizam uma variância local, como mostrado na seguinte equação:

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} g_{k,l}(I(i+k,j+l)\mu(i+k,j+l))^2}. \tag{3.15}$$

Os coeficientes MSCN são ajustados a uma GGD utilizando a equação 3.1. O BRISQUE expande seus atributos calculando um segundo conjunto de atributos a partir dos coeficientes MSCN já calculados. Nele é feita a multiplicação entre coeficientes de pixel vizinhos. A multiplicação é feita em quatro direções, conforme as seguintes equações:

$$H(i,j) = I_t(i,j) \cdot I_{t+1}(i,j+1), \tag{3.16}$$

$$V(i,j) = I_t(i,j) \cdot I_{t+1}(i+1,j), \tag{3.17}$$

$$D1(i,j) = I_t(i,j) \cdot I_{t+1}(i+1,j+1), \tag{3.18}$$

$$D2(i,j) = I_t(i,j) \cdot I_{t+1}(i+1,j-1). \tag{3.19}$$

O conjunto de atributos compostos por coeficientes MSCN secundários não segue uma distribuição Gaussiana Generalizada como os MSCN primários, sendo necessário fazer o ajuste da curva com uma distribuição Gaussiana Generalizada Assimétrica, AGGD. A distribuição AGGD possui duas variâncias, uma para cada lado da distribuição, que são denominadas variância esquerda e variância direita da distribuição. A fórmula de ajuste da distribuição sofre algumas alterações para acomodar a assimetria entre os lados da distribuição, sendo

$$f(x;\nu,\sigma_l^2,\sigma_r^2) = \begin{cases} \frac{\nu}{(\beta_l+\beta_r)\Gamma(1/\nu)} \cdot \exp\left[-\left(\frac{|x|}{\beta_l}\right)^{\nu}\right], & x < 0 \\ \\ \frac{\nu}{(\beta_l+\beta_r)\Gamma(1/\nu)} \exp\left[-\left(\frac{|x|}{\beta_r}\right)^{\nu}\right], & x \geq 0 \end{cases} \tag{3.20}$$

onde

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}}, \tag{3.21}$$

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}}, \tag{3.22}$$

$$\eta = (\beta_r - \beta_l)\frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})}, \tag{3.23}$$

onde ν e η são os parâmetros responsáveis pelo controle do formato da distribuição, σ_l^2 é variância do lado esquerdo da distribuição, σ_r^2 é a variância a direita da distribuição e $\Gamma(.)$ é a função gamma, já vista na Eq. 3.4.

Calculando todos os parâmetros do BRISQUE descritos acima, obtendo um total de 18 parâmetros, onde 2 são oriundos do ajuste dos coeficientes MSCN à distribuição Gaussiana generalizada e 16 são oriundos do ajuste dos 4 produtos de coeficientes MSCN à distribuição Gaussiana generalizada assimétrica. Esse processo é realizado em duas diferentes escalas dos quadros de vídeo: a escala original e uma escala reduzida obtida utilizando um filtro passa baixa e um escalonamento de 2 vezes. Esse processo é comum em algoritmos de qualidade por reproduzir um comportamento do sistema visual humano e tem mostrado bons desempenhos quando comparado ao uso de apenas uma escala [34].

3.3 Bases de dados

Durante o treinamento e o teste da NAVE, foram utilizadas 2 bases de dados. Procurou-se testar a métrica em múltiplas bases com diferentes degradações para que se pudesse ter um melhor entendimento do seu comportamento frente a degradações. Cada uma das bases utilizadas é descrita a seguir

3.3.1 Base de Dados Qualidade Audiovisual UnB 2018 (Experimento 1)

A Qualidade Audiovisual UnB 2018 (Experimento 1) [35] é a principal base de dados utilizada nesse trabalho. A base consiste em 3 experimentos com combinações de degradações de áudio e vídeo. O Experimento 1 contém apenas conteúdo de vídeo, o Experimento 2 possui apenas conteúdo de áudio e o Experimento 3 possui degradações de áudio e vídeo. Neste trabalho, utilizamos apenas o Experimento 1 cujo conteúdo é formado por 60 sequências de vídeo originais. A resolução espacial dos vídeos é de 1280×720, sua resolução temporal é de 30 quadros por segundo e o formato do espaço de cores é 4:2:0. A Tabela 3.1 mostra informações das sequências. Exemplos de quadros de alguns vídeos podem ser vistos na Figura 3.3.

A base de dados Experimento1 possuem degradações temporais e espacias. As degradações presentes nas sequências degradadas a partir das 60 sequências originais são compressão, perda de pacote e congelamento de quadros. São utilizados 2 algoritmos de compressão, 5 taxas de perda de pacotes e 5 cenários de congelamento de quadros na formação da base. Os algoritmos de compressão utilizado são o H.264/MPEG-4 Advance Video Coding (AVC) e o H.265 High Efficiency Video Coding (HEVC) [5] [6]. Ambos os compressores são os padrões mais utilizados atualmente, sendo o segundo uma geração mais recente do primeiro. Para cada um dos algoritmos, são utilizadas 4 taxas de bit, que podem ser vistos na Tabela 3.2.

A perda de pacote é um tipo de degradação que ocorre quando uma parcela dos pacotes enviado é perdido na transmissão. Essa perda pode causar diferentes degradações nos quadros, como *flickering* e blocagem. No base de dados Experimento 1, são utilizados 5 valores de taxa de perda de pacotes. Os valores recriam, juntamente com o codificador de vídeo, cenários reais de transmissão de vídeos via *streaming* [36] [37]. Os valores de perda de pacotes são: 1%, 3%, 5%, 8% e 10%. Tais valores foram combinados com os codificador de vídeo para criar 5 cenários que

Conteúdo	Sequência	Resolução espacial	Resolução temporal
Guy Sleeping	v01	1920 × 1080	30
Flamenco	v02, v60	1920 × 1080	30
Big Buck Bunny	v03, v04, v57	1920 × 1080	30
Elephant	v05, v06, v59	1920 × 1080	30
France Tourism	v07, v19, v53	1920 × 1080	30
WomanDay	v08, v18, v23, v45	1920 × 1080	30
Taiwan	v09, v13, v24, v32	1920 × 1080	30
Barca vs Atletic	v10, v17, v27, v40, v50	1920 × 1080	30
FootMusic	v11, v55	1920 × 1080	30
Atlanta Betline	v12, v30, v36, v37	1920 × 1080	30
Netflix El Fuente	v14, v26, v29, v38 v46, v49, v51, v54	1920 × 1080	30
Box interview NTIA	v15, v21, v58	1920 × 1080	30
Honey Bees	v16, v35	1920 × 1080	30
Kenpo Strikes NTIA	v20, v43	1920 × 1080	30
Taipei Fireworks	v22, v44	1920 × 1080	30
Old Town Car NTIA	v25	1920 × 1080	30
NTIA Violin	v28, v47	1920 × 1080	30
Puppies	v31, v48	1920 × 1080	30
Big Green Rabbit	v33	1920 × 1080	30
Movie Trailer Sintel	v34	1920 × 1080	30
Landscape Fast	v39	1920 × 1080	30
FoxBird	v41	1920 × 1080	30
Fishing Florida	v42, v56	1920 × 1080	30
Food	v52	1920 × 1080	30

Tabela 3.1: Informações dos vídeos presentes na base de dados de Qualidade Audiovisual UnB 2018 (Experimento1).

estão apresentados na Tabela 3.3.

Por último, a base de dados Experimento1 possui 5 cenários de congelamento de quadros. Os 5 cenários diferem quanto ao número, posição e duração dos eventos. Os vídeos podem conter entre 1 e 3 eventos de congelamento que podem estar distribuídos em 3 posições do vídeo. As possíveis posições são o início do vídeo, 1/3 da duração do vídeo e 2/3 da duração do vídeo. Os eventos são posicionados dessa forma para modelar os seguintes cenários de congelamento: evento de carregamento antes do vídeo ser tocado, evento de congelamento na primeira metade do vídeo e congelamento na segunda metade do vídeo. Esses 3 cenários são importantes porque cobrem cenários realistas, além de permitir o estudo do Efeito de Recência do evento de congelamento na experiência do espectador. A duração dos eventos pode ser igual a 1, 2 e 4 segundos. A Tabela 3.4 contém a duração, comprimento e as posições dos eventos de congelamento.

Em suma, a base de dados de Qualidade Audio Visual UnB 2018 (Experimento 1) possui 5 cenários de compressão e perda de pacotes, 5 cenários de compressão e congelamento de quadros e 2 cenários com vídeos codificados com uma taxa de bits elevada. Esses dois cenários são utilizados

Taxas	H.264	H.265
T1	500 kb	200 kb
T2	800 kb	400 kb
T3	2 Mb	1 Mb
T4	16 Mb	8 Mb

Tabela 3.2: Valores das taxas de bits para cada um dos Codecs de vídeo.

Cenário	Codec	Taxa de bits	TPP
HRC1	H.264	500 kb/s	10%
HRC2	H.265	400 kb/s	8%
HRC3	H.264	2000 kb/s	5%
HRC4	H.265	1000 kb/s	3%
HRC5	H.265	8000 kb/s	1%

Tabela 3.3: Cenários de degradação da base de dados Experimento 1 com seus respectivos codecs, taxas de bits e taxa de perda de pacotes (TPP).

como referência (âncora - ANC) de qualidade de para cada um dos codecs testados. O cenário ANC1 é codificado pelo H.264 com uma taxa 64Mb/s e o cenário ANC2 é codificado em HECV com taxa de 32Mb/s.

3.3.2 LIVE-Netflix-II

O segundo banco de dados utilizado neste trabalho é a LIVE-Netflix-II [38]. Essa base de dados foi criada como uma tentativa de modelar diferentes configurações de redes de transmissão de vídeos. Nela, são utilizados 15 vídeos originais que são codificados, transmitidos por rede simulada e recebidos por um cliente. O cenário utiliza quatro tipos de algoritmos de adaptação. No total, são geradas 420 sequências de vídeos que possuem resolução espacial de 1920×1080, resolução temporal de 24 quadros por segundo e formato do espaço de cores 4:2:0. As 7 redes simuladas nessa base de dados abrangem uma variedade de cenários com diferentes fatores limitantes, como por exemplo, diferentes valores médios de banda disponível e volatilidade da largura de banda

Cenário	Codec	Taxa de bit	Eventos	Posição	Duração
HRC10	H.264	16000kb/s	1	2	2
HRC9	H.264	2000kb/s	2	1,3	1,3
HRC8	H.265	1000kb/s	2	2,3	2,2
HRC7	H.264	800kb/s	3	1,2,3	2,2,2
HRC6	H.265	200kb/s	3	1,2,3	3,3,2

Tabela 3.4: Parâmetros de cada cenário de congelamento, onde a duração dos eventos está ordenada de acordo com a posição.

disponível. Essas redes são combinadas com 4 tipos de algoritmos adaptativos (*buffer-based*, *rate-based*, *quality-based* e *oracle quality-based*) de forma a criar 28 versões do 15 conteúdos originais.

Quando comparada à base da UnB, essa base possui semelhanças quanto ao conteúdo das degradações. Ambas possuem degradações oriundas das condições da rede de transmissão, porém com algumas diferenças. Enquanto a LIVE-Netflix-II possui efeitos de *rebuffering*, o Experimento 1 possui degradação de congelamento de quadro. Os eventos de *rebuffering* possuem um sinalizador de carregamento enquanto o vídeo está pausado, ao passo que o Experimento 1 não possui tal sinalizador e o vídeo permanece apenas com o quadro pausado. É esperado que esses eventos possuam diferenças desprezíveis na qualidade de experiência do espectador, porém não se sabe como os atributos irão responder às diferenças visuais dos dois cenários. Neste trabalho, a base de dados LIVE-Netflix-II foi utilizada para testar a capacidade de generalização da métrica em uma base com degradações semelhantes às degradações da base de dados Qualidade Audiovisual UnB 2018 (Experimento 1).

Conteúdo	**Resolução espacial**	**Resolução temporal**
Air Show	1920×1080	30
Asian Fusion	1920×1080	30
Chimera1102347	1920×1080	30
Chimera1102353	1920×1080	30
CosmosLaundromat	1920×1080	30
ElFuenteDance	1920×1080	30
ElFuenteMask	1920×1080	30
GTA	1920×1080	30
MeridianConversation	1920×1080	30
MeridianDriving	1920×1080	30
Skateboarding	1920×1080	30
Soccer	1920×1080	30
Sparks	1920×1080	30
TearsOfSteelRobot	1920×1080	30
TearsOfSteelStatic	1920×1080	30

Tabela 3.5: Informações dos vídeos presentes na base de dados LIVE-Netflix-II.

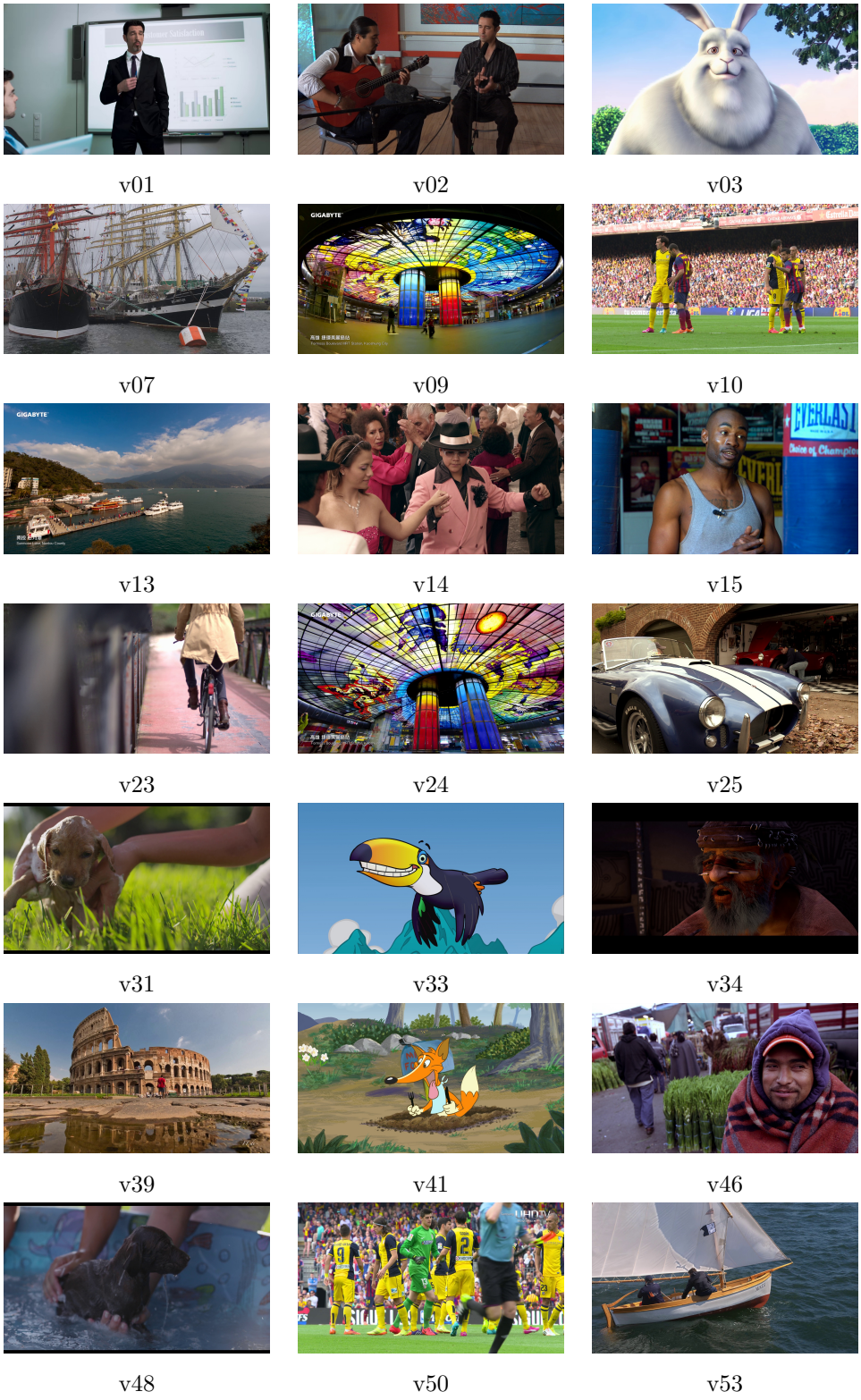


Figura 3.3: Exemplos de quadros de vídeos presentes na base de dados Experimento1 Unb Audiovisual quality database.

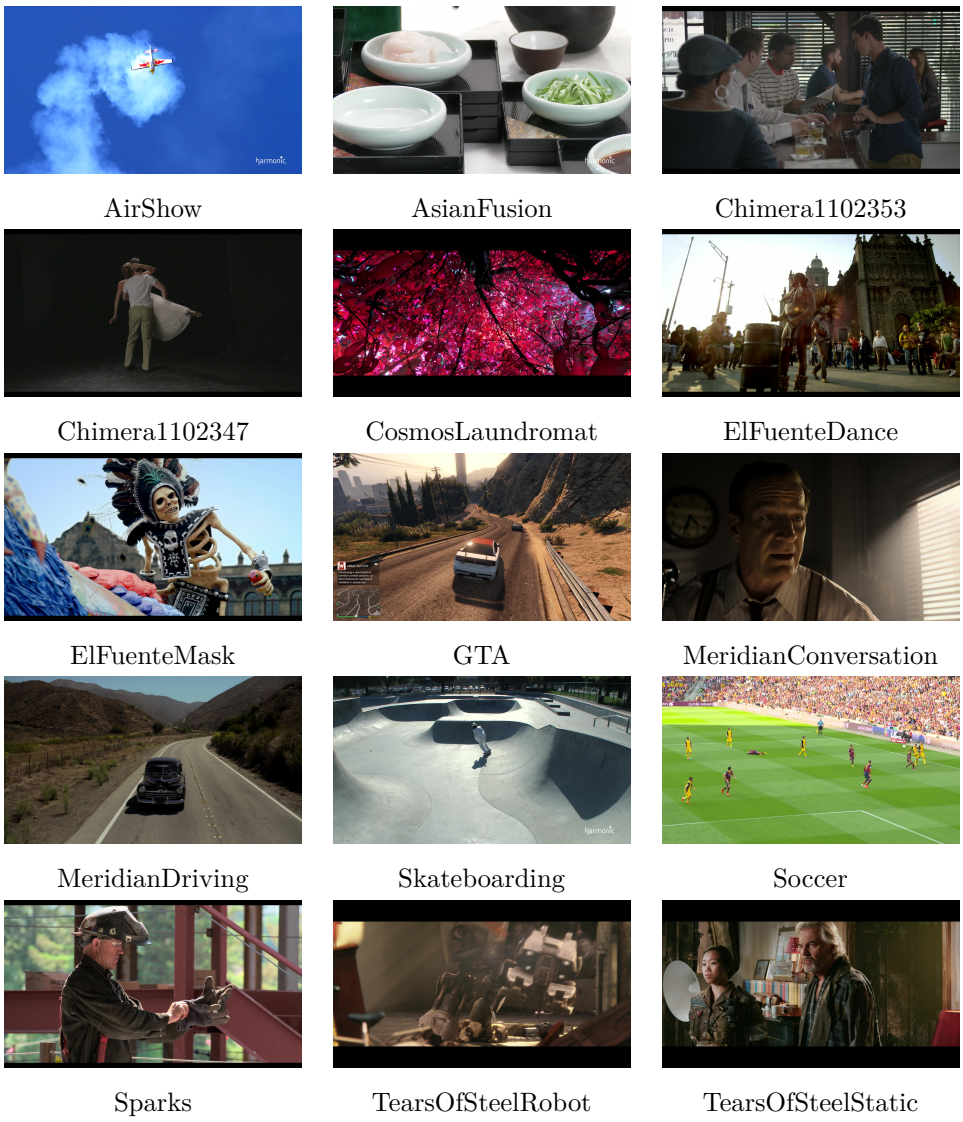


Figura 3.4: Exemplos de quadros de vídeos presentes na base de dados LIVE-Netflix-II.

Capítulo 4

Resultados Experimentais

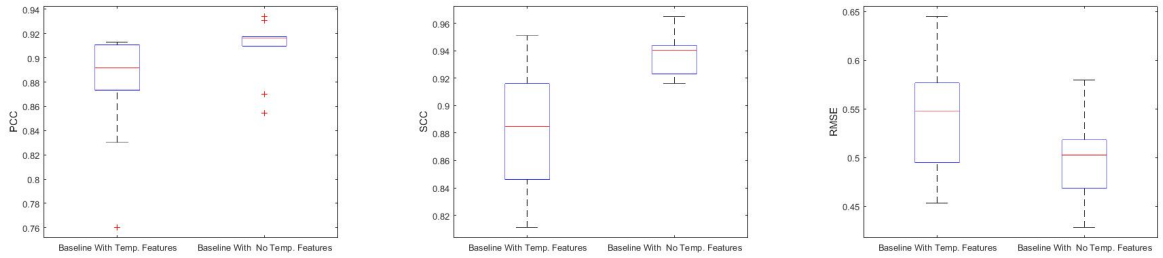
Neste capítulo, são apresentados os experimentos realizados. Os resultados obtidos pelas arquiteturas escolhidas também são apresentados e discutidos. Primeiramente, são apresentados os experimentos preliminares que motivaram a procura de novos atributos temporais para a métrica proposta neste trabalho, que denominamos **NAVEv2**. Em seguida, são realizados os testes dos novos candidatos a atributos, no qual, procura-se encontrar o melhor conjunto de novos atributos para trabalhar junto aos atributos DIIVINE. Após escolher o novo conjunto de atributos, a métrica é testada e analisada na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1) quanto ao seu desempenho frente a NAVE original e frente às degradações temporais. Por último, é feito o teste e a análise de desempenho da nova configuração da NAVE em uma segunda base de dados, a LIVE-Netflix-II.

4.1 Ablação Temporal

Em uma análise preliminar, investigou-se a contribuição dos hiper-parâmetros e os atributos utilizados na NAVE. Foi realizado uma análise de ablação nos atributos da métrica NAVE e descobriu-se que os atributos temporais utilizados na arquitetura original não contribuíam positivamente com a avaliação da qualidade.

O teste foi realizado na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1) que foi dividida em 10 partições. O treinamento e teste da métrica foi feito utilizando o esquema de *k-fold* nas 10 partições. Após cada teste, foi calculado o coeficiente de correlação de Pearson (PCC), o coeficiente de correlação de Spearman (SCC) e o erro médio quadrático (RMSE) da avaliação de qualidade da métrica.

Na Figura 4.1, é possível ver os coeficientes de correlação para dois cenários. No primeiro cenário, a NAVE foi treinada utilizando apenas informação espacial composta pelos atributos da métrica DIIVINE, enquanto, no segundo cenário, foram utilizadas informações espaciais e temporais, os atributos da métrica DIIVINE, a TI e a SI. Na Figura 4.1 (a), observa-se que o PCC para o cenário sem os atributos temporais apresenta uma distribuição mais fechada e uma mediana superior ao cenário com os atributos temporais. Um comportamento semelhante ocorre



(a) PCC (b) SCC (c) RMSE

Figura 4.1: Diagrama de caixa para o PCC, SCC e o RMSE para a NAVE treinada e testada, respectivamente, com os atributos DIIVINE e com o conjunto dos atributos DIIVINE, TI e SI.

para o SCC e RMSE, nas Figuras 4.1 (b) e (c), em que a distribuição da arquitetura apenas com atributos espaciais possui desempenho superior do que a arquitetura com atributos espaciais e temporais.

Esse resultado é inesperado uma vez que os atributos temporais deveriam auxiliar na detecção de degradações temporais, como perda de pacotes e congelamento de quadros. Todavia, as medidas TI e SI tiveram uma contribuição negativa na avaliação de qualidade dos vídeos. Outro ponto observado foi o bom desempenho dos atributos da métrica DIIVINE na detecção de degradações temporais. Não era esperado que os atributos espaciais apresentassem um bom resultado na avaliação de qualidade de degradações temporais.

4.2 Testes na Base de Qualidade Audiovisual UnB 2018 (Experimento 1)

O primeiro experimento realizado testou as diferentes combinações de conjuntos de atributos na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1). Os conjuntos de atributos foram combinados em 8 grupos, que são apresentados na Tabela 4.1. Todos os conjuntos possuem os atributos DIIVINE como base, exceto, os conjuntos 2 e 8. Devido ao resultado positivo apresentado anteriormente pelos atributos DIIVINE, eles foram mantidos na maioria dos conjuntos de atributos.

4.2.1 Resultados dos Conjuntos de Atributos

Na Tabela 4.2 são apresentados os valores de PCC, SCC e RMSE para os conjuntos de atributos, onde cada valor reportado é a média dos testes feitos em cada uma das 10 partições da base de dados. Os coeficientes estão agrupados por conjunto de atributos e separados. As colunas da tabela mostram os resultados para os diversos tipos de degradações presentes nas base de dados. A primeira coluna apresenta os valores para os vídeos degradados apenas com perda de pacotes, enquanto a segunda coluna apresenta os resultados para os vídeos com congelamento de quadros. Por fim, a última coluna apresenta os coeficientes para todos os vídeos da base de dados. Em cada

Conjunto de atributos	Métrica original
1	DIIVINE
2	BRISQUE + SinnoP
3	DIIVINE + BRISQUE
4	DIIVINE + SinnoM
5	DIIVINE + SinnoP
6	DIIVINE + BRISQUE + SinnoM
7	DIIVINE + BRISQUE + SinnoP
8	SinnoP

Tabela 4.1: Tabela contendo os conjuntos de atributos usados no treinamento e teste da NAVE na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1).

coluna, os melhores valores dos coeficientes estão destacados em negrito.

É possível notar que o conjunto 1, contendo apenas os atributos DIIVINE, obteve uma alta correlação com a Pontuação de Opinião Média, em inglês *Mean Opinon Score* (MOS). Os valores PCC e SCC foram elevados para os dois tipos de degradação e, por consequência, para o cenário com todos os vídeos. Comparando o conjunto 1 aos conjuntos 4 e 5, é possível notar a influência dos atributos Sinno. No conjunto 5, temos um decréscimo da correlação e um aumento do valor RMSE. A maior queda está nos valores do SCC, o que indica uma redução da monotonicidade entre os valores de qualidade preditos pelo modelo e os valores de MOS. Além disso, existe uma piora nos valores gerais e nos valores do cenário com congelamento de quadros. No conjunto 4, observa-se um comportamento diferente do conjunto 5. Não existe uma queda acentuada dos valores de correlação para a perda de pacotes e uma melhoria dos valores para os vídeos com congelamento de quadros. Neste cenário, o PCC se manteve com um valor muito semelhante ao do conjunto 1, ao passo que o valor de SCC subiu ligeiramente e o do RMSE reduziu.

Os resultados da Tabela 4.2 para os conjuntos 4 e 5 apontam que o maior número de atributos do conjunto SinnoP não apresenta melhora no desempenho se comparado ao conjunto SinnoM que possui menos atributos. A comparação entre os conjuntos SinnoM e SinnoP pode ser expandida para os conjuntos de atributos 6 e 7. O conjunto 6 apresenta os atributos espaciais DIIVINE e BRISQUE e os atributos temporais SinnoM enquanto o conjunto 7 apresenta os atributos SinnoM são substituídas pelos atributos SinnoP. O conjunto 6 apresenta os melhore valores de PCC e RMSE dentre todos os conjuntos para a degradação de perda de pacotes. No entanto, existe uma redução do SCC para este cenário. Para o cenário de congelamento de quadros, há uma grande redução do PCC no conjunto 6 se comparado aos conjuntos 1 e 4. Contudo, os valores do conjunto 6 são superiores aos valores de coeficientes para o conjunto 7 que contém os atributos SinnoP. O conjunto 7 obteve valores piores para dos 3 coeficientes apresentados na Tabela 4.2 para os dois tipos de degradação e para o caso geral. Após comparar os conjuntos 4 e 5 e os conjuntos 6 e 7, é possível notar que existe uma clara diferença entre o desempenho no que diz respeito às versões dos atributos Sinno. Os atributos SinnoM tiveram um desempenho consistentemente superior aos atributos SinnoP.

Os atributos BRISQUE foram o segundo conjunto de atributos testado neste trabalho. Nova-

Conjunto de atributos	Coeficiente	Perda de Pacotes	Congelamento	Todos
Conjunto de atributos 1	PCC	0.944	**0.910**	0.909
DIIVINE	SCC	**0.965**	0.920	**0.937**
	RMSE	0.426	0.477	0.457
Conjunto de atributos 2	PCC	0.944	0.763	0.833
BRISQUE + SinnoP	SCC	0.834	0.822	0.796
	RMSE	0.424	0.588	0.515
Conjunto de atributos 3	PCC	0.948	0.849	0.886
DIIVINE + BRISQUE	SCC	0.902	0.914	0.907
	RMSE	0.386	0.486	0.442
Conjunto de atributos 4	PCC	0.941	0.909	**0.905**
DIIVINE + SinnoM	SCC	0.937	**0.925**	0.915
	RMSE	0.424	**0.456**	0.444
Conjunto de atributos 5	PCC	0.934	0.889	0.894
DIIVINE + SinnoP	SCC	0.885	0.891	0.886
	RMSE	0.456	0.493	0.477
Conjunto de atributos 6	PCC	**0.951**	0.851	0.887
DIIVINE + BRISQUE + SinnoM	SCC	0.942	0.920	0.925
	RMSE	**0.370**	0.489	**0.436**
Conjunto de atributos 7	PCC	0.949	0.834	0.878
DIIVINE + BRISQUE + SinnoP	SCC	0.914	0.897	0.904
	RMSE	0.388	0.507	0.454
Conjunto de atributos 8	PCC	0.834	0.817	0.807
SinnoP	SCC	0.805	0.765	0.761
	RMSE	0.591	0.724	0.662

Tabela 4.2: Média dos valores dos PCC, SCC e RMSE para os testes realizados na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1).

mente, considerou-se o conjunto 1 como a base de comparação. É possível notar que o conjunto 3 que contém os atributos DIIVINE e BRISQUE, possui um desempenho geral pior do que o conjunto 1. A adição dos atributos BRISQUE reduziu o SCC para o cenário de perda de pacotes ao passo que o PCC se manteve relativamente constante com um pequeno aumento. No cenário de congelamento de quadros, houve uma redução do PCC enquanto que o SCC se manteve relativamente constante. O segundo cenário que deve ser visto é a comparação entre os conjuntos 4 e 6, que incluem os atributos DIIVINE e SinnoM. É possível notar que a média dos valores do PCC para o cenário de perda de pacotes é maior para o conjunto 6 que contém os atributos BRISQUE se comparado ao conjunto 4. Este é o maior valor de PCC obtido entre todos os conjuntos de atributos. É possível notar nos conjuntos 3 e 6 que os atributos BRISQUE aumentam a linearidade da MOS predita pelo modelo. No entanto, esses mesmos atributos possuem um efeito negativo na SCC, que é reduzida nos conjuntos 3 e 6 para o cenário com perda de pacotes. O conjunto 6 obteve o melhor valor de RMSE dentre todos os conjuntos de atributos. Observando o cenário de congelamento de quadros, é possível notar que existe uma queda grande do PCC para os conjuntos 4 e 6. Ocorre uma leve queda no SCC, porém o valor é igual ao valor obtido para o conjunto 1, que é o conjunto base. O valor de RMSE é superior ao valor obtido pelo conjunto 4.

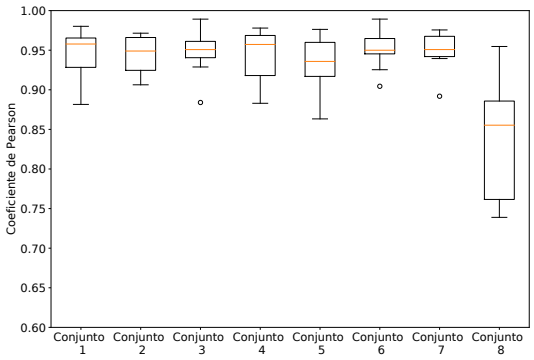
Comparando os cenários 5 e o cenário 7, percebemos que existe uma melhora nos valores de PCC e SCC para o cenário de perda de pacotes. Por outro lado, o valor do PCC cai para o cenário de congelamento de quadro. A queda dos coeficientes nos cenários de congelamento de

quadros é inferior ao aumento presente nos cenários de perda de pacotes, o que gera uma melhora de desempenho ao se adicionar os atributos BRISQUE ao conjunto de atributos. Os resultados obtidos na comparação dos conjuntos 4 e 6 e dos conjuntos 5 e 7 mostram um pior de desempenho da métrica quando os atributos BRISQUE são utilizados em conjunto com as SinnoM. O mesmo não é observado quando estes atributos são utilizados em conjunto com os atributos SinnoP. Não houve um cenário onde a adição dos atributos BRISQUE levaram a um resultado positivo, com exceção do cenário de perda de pacotes para o conjunto 6. Uma conclusão semelhante pode ser feita para os atributos SinnoP. Em nenhum dos casos em que ela foi adicionada, houve melhora do desempenho médio da métricas.

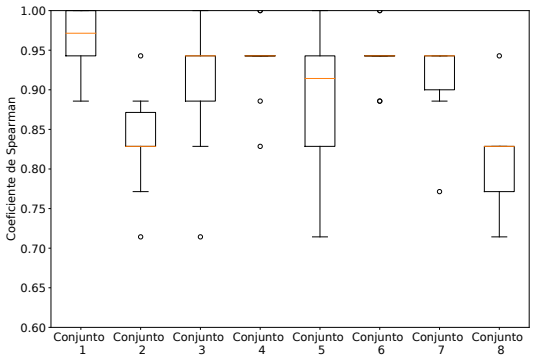
Na Tabela 4.2 são apresentados os valores médios para os coeficientes de correlação para o desempenho da NAVE. No entanto, a análise apenas os valores médios não representam uma ideia muito clara do comportamento. Desta forma, realizamos uma análise adicional levando em conta a distribuição dos valores obtidos no diferentes *k-folds*. Esta análise permite distinguir resultados nos quais os valores médios são similares e as distribuições diferentes são diferentes. As Figuras 4.2 (a), (b) e (c) apresentam os diagramas de caixa para os coeficientes de correlação no cenário com degradação de perda de pacotes. Na Figura 4.2 (a), é possível notar que neste cenário existe uma diferença clara entre o conjunto 8 e os demais conjuntos, onde o conjunto 8 apresenta uma distribuição com valores inferiores aos valores obtidos para os demais cenários. Outra distinção que pode ser feita é entre os conjuntos 3, 6 e 7 e os demais conjuntos. Estes conjuntos possuem distribuição mais compacta, onde os valores inferiores são superiores aos valores inferiores das outras distribuições. O ponto comum entre esses 3 conjuntos é a presença dos atributos BRISQUE, indicando uma correlação entre a informação capturada por estes atributos e a degradação de perda de pacotes.

A Figura 4.2 (b) mostra o diagrama de caixa do SCC para os diferentes conjuntos de atributos para o cenário de perda de pacotes. De forma semelhante ao diagrama do PCC, o conjunto 8 possui valores inferiores aos valores obtidos para os demais cenários. Os conjuntos 3, 5 e 7 possuem valores de coeficientes de correlação inferiores aos valores obtidos por outros conjuntos de atributos. Os conjuntos 4 e 6 não tiveram uma distribuição tão compacta, enquanto que o conjunto 1 possui a distribuição com os valores mais elevados dentre todas as distribuições. As distribuições mais compactas não possuem necessariamente os atributos BRISQUE. Ao contrário, duas das distribuições não possuem tais atributos. Sendo assim, não existe uma ligação direta entre os atributos BRISQUE e o aumento de desempenho da métrica no cenário de perda de pacotes.

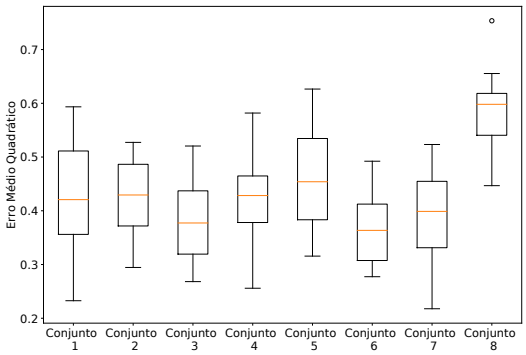
As Figuras 4.3 (a), (b) e (c) apresentam os diagramas de caixa dos cenários contendo congelamento de quadros para os testes realizados. Na Figura 4.3 (a), pode se notar resultados com desempenhos variados que podem estar associados com diferentes agrupamentos. Primeiramente, os conjuntos 1 e 4 que obtiveram os valores mais altos para os máximos, as medianas e os mínimos dentre todas as distribuições. Em seguida, temos os conjuntos 3, 5 e 7 que obtiveram valores medianos de desempenho. Finalmente, os conjuntos 2, 6 e 8 obtiveram resultados inferiores a todos os conjuntos supracitados. Neste último agrupamento, o conjunto 6 se destaca pela grande dispersão da distribuição dos valores, onde o máximo da distribuição é igual aos valores superiores



(a) PCC



(b) SCC

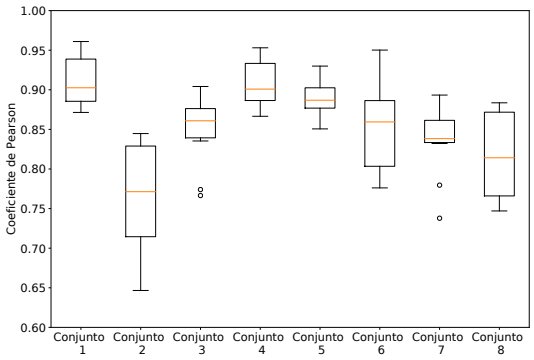


(c) RMSE

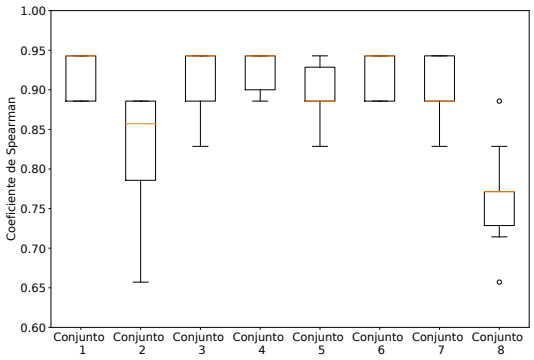
Figura 4.2: Diagrama de caixa do PCC, SCC e RMSE para os diferentes conjuntos de atributos no teste da NAVE para os cenários contendo perda de pacotes.

obtidos pelos conjuntos 1 e 4. Porém, o seu mínimo e a sua mediana são os valores mais baixos dentre todos os conjuntos. É possível notar que os conjuntos 1, 4 e 6, que contém os atributos DIIVINE e, possivelmente, atributos SinnoM, possuem os valores máximos mais altos entre todas as distribuições. Comparando os conjuntos 5 e 7 aos conjuntos 4 e 6, é possível notar que os atributos SinnoM obtiveram resultados superiores as SinnoP. Por último, observa-se que os atributos BRISQUE causam uma piora do desempenho em todos os cenários nos quais elas foram adicionadas.

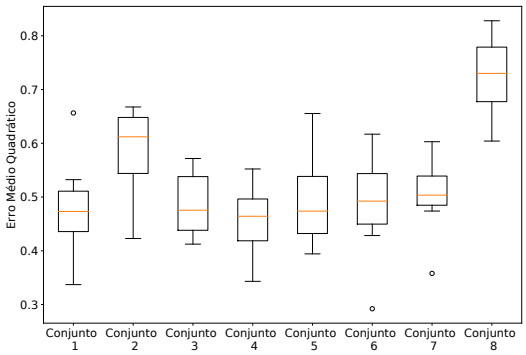
Na Figura 4.3 (b), é possível notar que existe um valor máximo do qual nenhum dos conjuntos conseguiu passar e que houve distribuições de caixa com formatos anormais para grande parte dos conjuntos. Existe uma sobreposição dos valores de máximo e terceiro percentil e uma sobreposição dos valores de mínimo e primeiro percentil. De forma semelhante ao ocorrido para os PCC, é possível notar que os conjuntos contendo os atributos SinnoP obtiveram resultados inferiores ao conjunto 1. Os conjuntos 5 e 7 obtiveram medianas e valores mínimos inferiores aos do conjunto 1, que contém apenas os atributos DIIVINE.



(a) PCC



(b) SCC

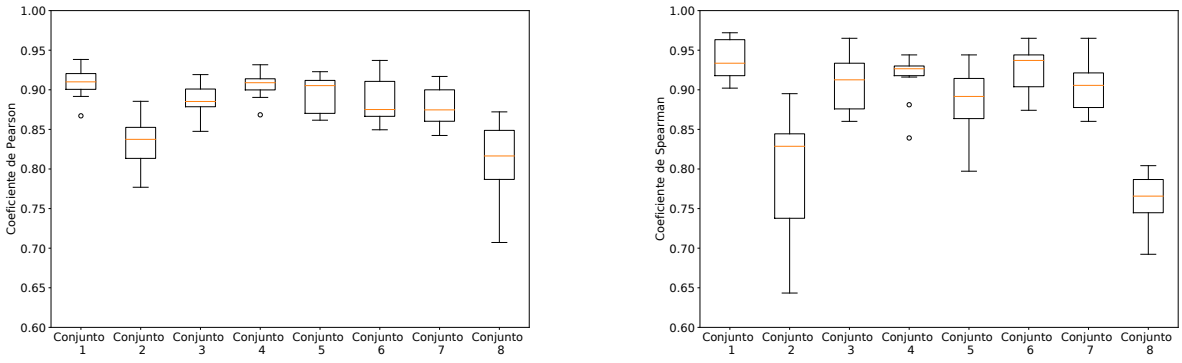


(c) RMSE

Figura 4.3: Diagrama de caixa do PCC, SCC e RMSE para os diferentes conjuntos de atributos no teste da NAVE para os cenários contendo congelamento de quadros.

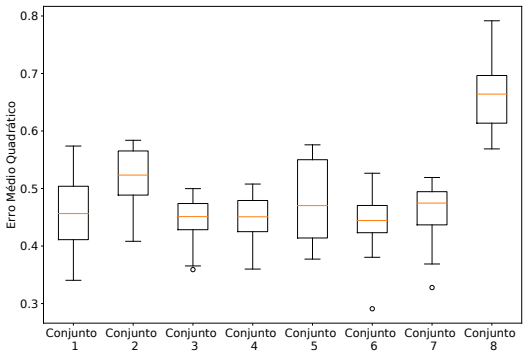
Os conjuntos que contêm os atributos SinnoM possuem resultados similares aos encontrados para o conjunto 1. A distribuição do conjunto 6 é muito semelhante à distribuição do conjunto 1, enquanto que a distribuição do conjunto 4 possui resultado superior ao conjunto 1. É possível observar que o conjunto 4 possui o valor do seu primeiro percentil superior ao mesmo valor do conjunto 1. Isto indica que existe uma concentração dos valores da distribuição do conjunto 4 em valores superiores se comparado ao conjunto 1. Na Figura 4.3 (c), observa-se um comportamento semelhante ao encontrado nos diagramas de caixa do PCC e SCC, nas quais, os atributos SinnoM possuem resultados superiores ou iguais ao conjunto 1. No entanto, os atributos SinnoP e BRISQUE influenciam negativamente o desempenho da métrica.

As Figuras 4.4 (a), (b) e (c) apresentam os diagramas de caixa para o NAVE testada em todos os cenários de degradação da base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1). Pode-se notar que os conjuntos 1 e 4 se destacam com coeficientes de correlação superiores aos demais conjuntos. Para ambos PCC e SCC, a distribuição está bem mais concentrada ao redor da mediana do que nos demais conjuntos. É possível notar também que os conjuntos 2 e 8 tiveram



(a) PCC

(b) SCC



(c) RMSE

Figura 4.4: Diagrama de caixa do PCC, SCC e RMSE do desempenho dos diferentes conjuntos de atributos no teste da NAVE para todas as degradações.

as piores distribuição de coeficientes.

Os conjuntos 4 e 6 que contêm os atributos SinnoM obtiveram os maiores valores de correlação com a MOS dentre os conjuntos de atributos juntamente com o conjunto 1. O conjunto 4 obteve uma distribuição similar ao conjunto 1 para o PCC. A diferença entre os 2 conjuntos está na distribuição do SCC, onde o conjunto 4 obteve uma distribuição mais fechada e com mesmo valor médio se comparado ao conjunto 1. A diminuição da distribuição é um caso de melhora que não pode ser observado através dos valores da Tabela 4.2. O conjunto 6 que também contém os atributos SinnoM apresenta valores altos semelhantes ao conjunto 4 porém sua distribuição é mais dispersa. Essa dispersão pode ser atribuída aos atributos BRISQUE que é a diferença entre os dois conjuntos. É possível notar também que a distribuição do conjunto 6 é pior do que a distribuição do conjunto 1 que contém apenas atributos espaciais.

Os conjuntos contendo os atributos SinnoP não obtiveram resultados superiores ao conjunto 1. Em todos os cenários onde estes atributos foram adicionados não houve melhora se comparado aos conjuntos contendo os atributos SinnoM. Quando os dois tipos de atributos Sinno são comparados

isoladamente através dos conjuntos 4 e 5, pode se notar um espalhamento da distribuição para ambos os coeficientes de correlação sendo o espalhamento para o SCC mais severo do que para o PCC. É válido ressaltar que a piora para o SCC é severa e é o pior resultado dentre os conjuntos com contém os atributos DIIVINE. Os diagramas de caixa apontam a não existência de melhora caso sejam utilizados os atributos Sinno com os coeficientes de todo os *patches* do cada quadro do vídeo.

Comparando os conjuntos 6 e 7, nos quais os atributos BRISQUE são adicionadas, a diferença de desempenho não é tão discrepante. O desempenho dos atributos SinnoP continua pior do que o desempenho dos atributos SinnoM. No entanto, a diminuição de discrepância se deve aos atributos BRISQUE, uma vez que o conjunto 3, que contém os atributos DIIVINE e BRISQUE, possui uma distribuição bastante similar ao conjunto 7, que contém o conjunto 3 e os atributos SinnoP. A diferença da contribuição dos atributos SinnoP e BRISQUE se torna evidente nos valores de SCC, no qual a distribuição correspondente ao conjunto 5 é mais dispersa do que a distribuição correspondente ao conjunto 3.

Os atributos BRISQUE estão presentes nos conjuntos 2, 3, 6 e 7. No conjunto 2, onde eles foram combinados aos atributos SinnoP, houve uma clara piora dos valores do PCC e SCC. Quando os atributos BRISQUE são combinados com os atributos DIIVINE no conjunto 3, não há uma melhora na correlação. No entanto, devemos ressaltar que a queda de desempenho não foi tão severa quanto no conjunto 5, que contém os atributos SinnoP. Nos conjuntos 6 e 7, os atributos BRISQUE foram combinados respectivamente com os atributos dos conjuntos 4 e 5. No conjunto 7, não existe uma queda de desempenho quando comparado ao conjunto 5, porém existe uma queda em relação ao conjunto 1. Desta forma, é mais vantajoso utilizar apenas os atributos DIIVINE do que os atributos SinnoP e BRISQUE em conjunto. O conjunto 6 possui valores máximos da distribuição semelhantes aos valores dos conjuntos 1 e 4. Porém, o mesmo não pode ser observado para os valores inferiores da distribuição dos SCC. Considerando que os valores inferiores das distribuições correspondem aos vídeos mais difíceis de serem avaliados, pode-se dizer que os atributos BRISQUE reduzem a capacidade do modelo de detectar precisamente as degradações de compressão e degradações temporais.

4.2.2 Comparação com Outras Métricas

Com base nas diversas comparações entre os conjuntos de atributos apresentados anteriormente, o conjunto de atributos 4, que contém os atributos DIIVINE e SinnoM, se destacou pelos valores elevados de correlação tanto para cenários de perda de pacotes como para os cenários de congelamento de quadros. A contribuição dos atributos SinnoM é vista na distribuição compacta dos valores de correlação, com os valores inferiores das distribuições mais elevados dentre todos os conjuntos de atributos. Por estas razões, escolhemos o conjunto 4 para ser utilizado nas próximas comparações da NAVEv2 com outras métricas. A partir deste ponto, o nome NAVE é utilizado para descrever a arquitetura original da NAVE, que utiliza os atributos DIIVINE, SI e TI, enquanto que o nome NAVEv2 é a NAVE com os atributos DIIVINE e SinnoM. Comparamos a métrica NAVEv2 com as seguintes métricas:

- Métricas FR: SSIM [39] and PSNR;
- Métricas NR: DIIVINE [26], BIQI [40], NIQE [32], and BRISQUE [34], VIIDEO [16] e NAVE [2].

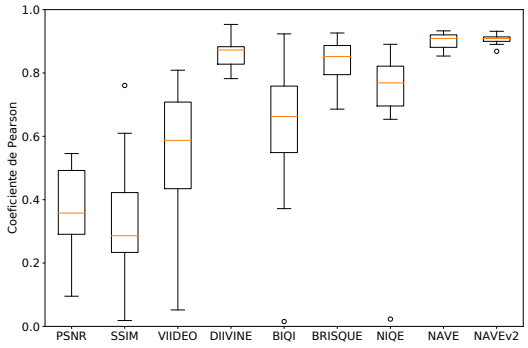
Todas as métricas foram treinadas e testadas utilizando um *k-fold* de 10 partições na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1).

A Tabela 4.3 mostra os resultados da comparação, ou seja, os valores médios das PCC, SCC e RMSE. Os resultados foram obtidos usando um *k-fold* para treinar e testar as métricas com referência e sem referência nos três cenários de degradação presentes na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1). Os cenários estão separados por degradação: (1) cenário com apenas perdas de pacotes; (2) cenário com apenas congelamentos de quadros; (3) cenário com todas as degradações. Na Tabela 4.3, os valores destacados em negrito correspondem aos maiores valores obtidos para aquela medida entre todas as métricas. É possível notar que a NAVEv2 obteve um resultado superior às demais métricas, assim superando a versão original da métrica NAVE. Para o cenário das perdas de pacotes, há uma melhora nos valores do PCC e do RMSE. Percebe-se também uma redução do RMSE para o cenário de perda de pacotes, enquanto que o SCC sofreu uma pequena diminuição. No cenário de congelamento de quadros, houve uma melhora consistente de todos os coeficientes de correlação analisados. No cenário com todas as degradações, houve melhora do PCC e do RMSE, enquanto os valores do SCC se mantiveram constantes.

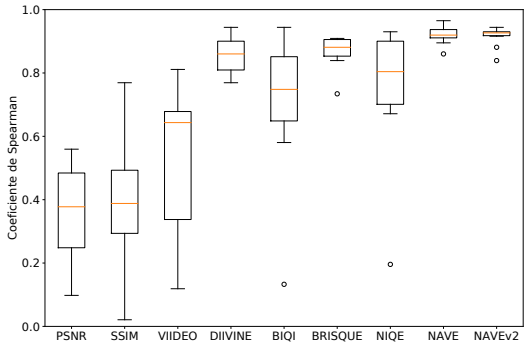
As Figuras 4.5 (a) e (b) apresentam respectivamente os diagramas de caixa para o PCC e SCC para as diferentes métricas. Dentre os resultados observados, é possível notar que grande parte das métricas de imagem e de vídeo obtiveram resultados com baixa correlação com a MOS. Dentre as métricas de imagens de melhor desempenho estão a DIIVINE e a NIQE que obtiveram correlações relativamente altas. Dentre essas, podemos destacar a métrica DIIVINE que obteve os resultados mais semelhantes dentre todas as métricas com exceção das versões da NAVE. É possível notar que o resultado da métrica DIIVINE se aproxima ao resultado obtido pela NAVE. Esta semelhança pode ser explicada pelo fato de ambas utilizarem conjuntos muito semelhantes de atributos, sendo que a diferença entre elas é as respectivas funções de classificação e a redução do número de atributos via autoencoders. É possível notar também que para o PCC houve um achatamento da distribuição e uma concentração dos valores em um valor próximo de $0,9$. Comparando os resultados já obtidos com os resultados da NAVEv2, é possível ver que houve uma concentração ainda maior da distribuição em torno da mediana e que os valores mais baixos obtidos pela distribuição são superiores aos valores mais baixos obtidos NAVE original para o PCC.

Através das Figura 4.5 (a) e (b), é possível notar que as mesmas métricas que obtiveram uma baixa correlação com MOS para o PCC também apresentaram uma baixa correlação para o SCC. É possível notar também que houve resultados similares entre as métricas DIIVINE, BRISQUE e a NAVE, onde a BRISQUE ficou entre as métricas com melhor desempenho. Novamente, é possível notar que a NAVEv2 tem um melhor desempenho com relação a NAVE tanto para o PCC como para o SCC.

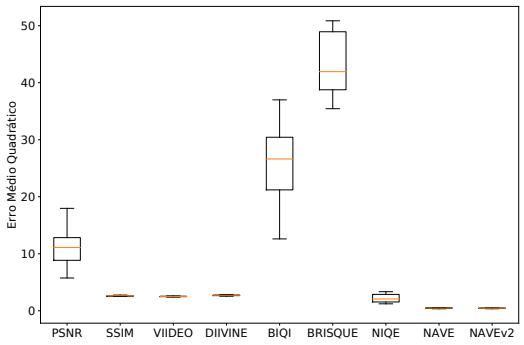
A melhora de desempenho pode ser vista principalmente na parte inferior das distribuições, onde os valores mais baixos da NAVEv2 foram superiores aos valores mais baixos da NAVE. Na Tabela 4.3, observa-se que a substituição dos atributos SI e TI pelos atributos SinnoM melhorou o desempenho da métricas para os dois cenários de degradação. A melhora mais significativa aconteceu para os vídeos com congelamento de quadros.



(a) PCC



(b) SCC



(c) RMSE

Figura 4.5: Diagrama de caixa do PCC, SCC e RMSE de diferentes métricas para todos os cenários do Experimento 1.

4.2.3 Degradações de compressão e degradações temporais

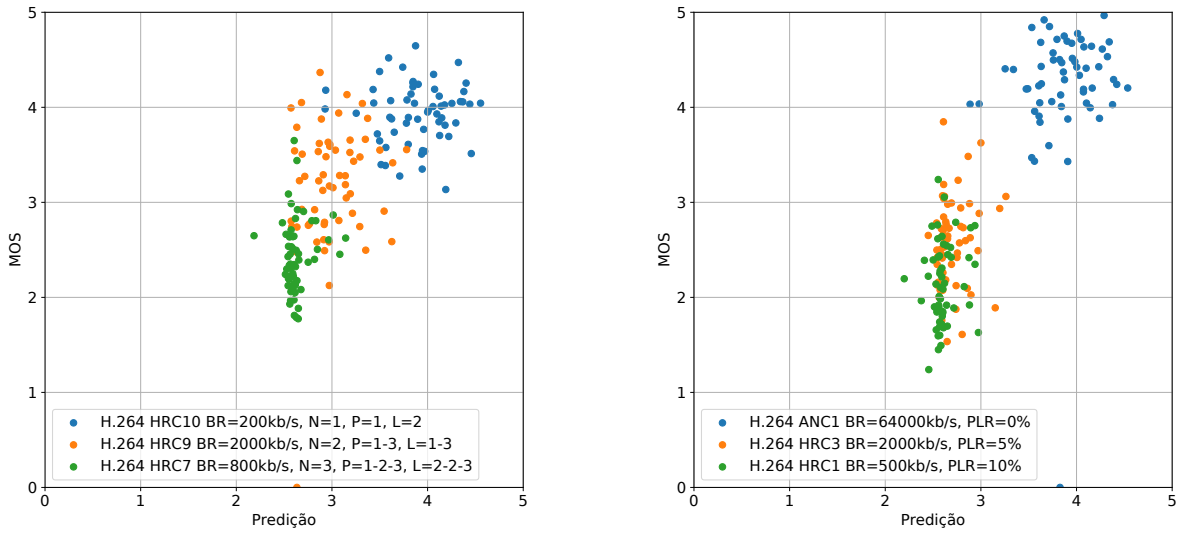
Através das análises anteriores, é possível notar que a NAVEv2 obteve uma melhora de desempenho com relação a métrica original. Porém, fica a dúvida se a métrica foi realmente capaz de detectar as degradações temporais uma vez que a base de dados utilizada não possui vídeos com as degradações isoladas. O objetivo da análise é verificar a capacidade da métrica de avaliar a qualidade dos vídeos, levando em consideração as degradações temporais. Uma das possibilidades seria a métrica não identificar corretamente ou não ser afetada pelas degradações tempo, sendo sensível apenas a diferenças de codificação.

Conjunto de features	Medida	Perda de Pacotes	Congelamento	Todos
PSNR	PCC	0.363	0.661	0.310
	SCC	0.434	0.600	0.314
	RMSE	13.623	8.664	11.436
SSIM	PCC	0.103	0.578	0.157
	SCC	0.125	0.577	0.163
	RMSE	2.364	2.852	2.621
DIIVINE	PCC	-0.889	-0.89	-0.858
	SCC	-0.862	-0.85	-0.855
	RMSE	2.515	2.881	2.705
VIIDEO	PCC	-0.578	-0.517	-0.542
	SCC	0.571	-0.440	-0.500
	RMSE	2.304	2.685	2.502
BIQI	PCC	-0.756	-0.576	-0.692
	SCC	-0.731	-0.640	-0.694
	RMSE	25.380	24.532	24.326
NIQE	PCC	-0.569	0.808	-0.702
	SCC	-0.634	0.828	-0.714
	RMSE	2.226	2.109	2.179
BRISQUE	PCC	-0.772	-0.887	-0.835
	SCC	-0.817	-0.908	-0.866
	RMSE	2.226	41.726	43.367
NAVE	PCC	0.933	0.895	0.896
	SCC	**0.942**	0.914	**0.917**
	RMSE	0.428	0.499	0.468
NAVEv2	PCC	**0.941**	**0.909**	**0.905**
	SCC	0.937	**0.925**	0.915
	RMSE	**0.424**	**0.456**	**0.444**

Tabela 4.3: Média dos valores dos PCC, SCC e RMSE para as diferentes métricas treinadas e testadas na base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1).

A Figura 4.6 apresenta o diagrama de dispersão da predição de qualidade da NAVEv2 versus MOS para os vídeos codificados com o H.264. Os vídeos foram separados de acordo com as degradações (congelamento de quadros e perda de pacotes). Na Figura 4.6 (a), é possível observar os casos com congelamento de quadros, onde o parâmetro N indica a quantidade de eventos de congelamento presente no vídeo, o parâmetro P indica a posição da degradação (começo, meio ou fim do vídeo) e o parâmetro L indica a duração do evento de congelamento (curta, média ou longa).

A Figura 4.6 (b) apresenta os vídeos degradados com perda de pacotes, com o parâmetro PLR representando a taxa de perda de pacotes. Nesta figura é possível notar que os casos HRC3 e HRC1 possuem nuvens dispersões que se sobrepõe, indicando que o modelo é incapaz de diferenciar os dois cenários. Os casos HRC3 e HRC1 estão em uma posição diferente do caso ANC1. Existe uma separação bem definida entre os casos com e sem degradação. Pode-se notar que a NAVE não consegue distinguir entre dois cenários de perda de pacotes quando estes cenários estão sendo codificados com o codec H.264.



(a) Congelamento de quadros

(b) Perda de Pacotes

Figura 4.6: Diagrama de dispersão do vídeos da base Experimento1 codificados em h264 e contendo congelamento de quadros e vídeos codificados em h264 e contendo perda de pacotes.

Nos cenários com degradação de congelamento de quadro da Figura 4.6(a), é possível notar que existe uma separação muito mais bem definida entre os três cenários de degradação. O cenário HRC10, que possui as degradações de congelamento menos severas, está claramente separado do caso HRC9 que possui mais compressão e um maior número de eventos de congelamento de quadro em posições e em comprimentos superiores ao caso anterior. Por último, temos o caso HRC7 que possui uma compressão mais elevada do que os casos anteriores, onde há o maior número de eventos de congelamento de quadro com as maiores durações.

Quando comparamos os cenários com congelamento de quadro e os cenários com perda de pacote, pode-se notar que os cenários com congelamento de quadro estão dispostos nas regiões mais próximas à diagonal do diagrama. Nos cenários de perdas de pacotes, é possível notar que os vídeos se encontram mais distantes da diagonal da Figura. A diagonal representa o cenário onde a estimativa de qualidade da NAVEv2 está igual a MOS de cada vídeo. Quanto mais próximo o ponto se encontra da diagonal, melhor foi o desempenho da métrica. Em especial, foram atribuídas incorretamente notas de qualidade entre 2 e 3 para vídeos do grupo HRC1 que possuem MOS inferior a 2. É possível notar que existe uma clara diferença entre a capacidade da métrica, dependendo do tipo de degradação que está acompanhando a compressão H.264.

A Figura 4.7 apresenta o diagrama de dispersão da predição de qualidade da NAVEv2 versus MOS para os vídeos codificados com o H.265. Nestas figuras, é possível notar que existe um comportamento oposto para as combinações entre as degradações temporais e a codificação H.265, se comparado às combinações com codificação H.264. Em vários vídeos codificados em H.265, NAVEv2 teve dificuldade para identificar e separar as degradações de congelamento de quadros do que para as degradações de perda de pacotes.

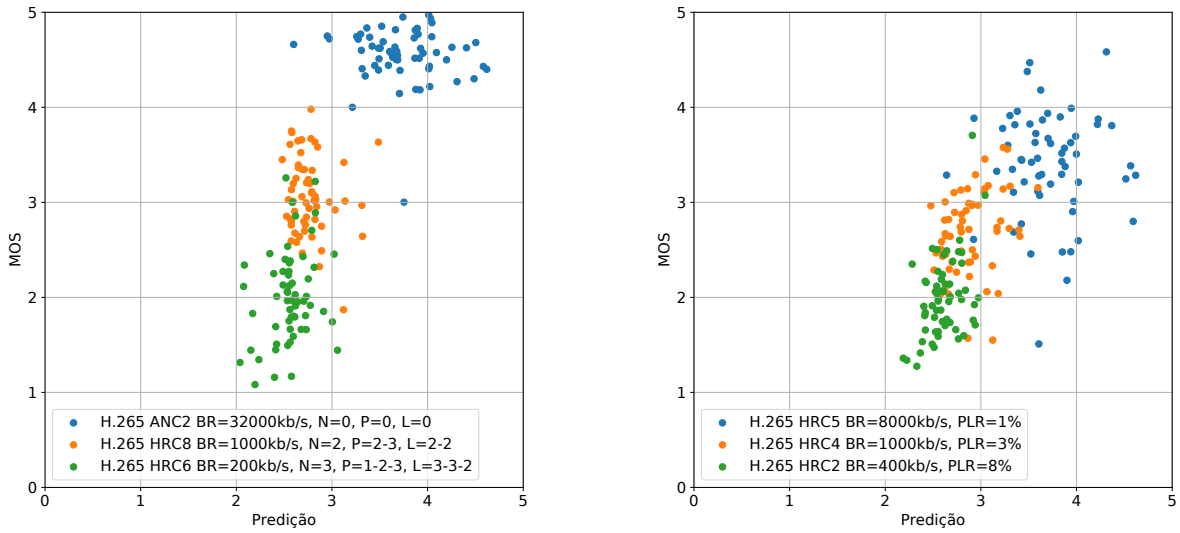
Na Figura 4.7 (a), é possível notar uma predição incorreta nos conjuntos HRC8 e HRC6. Existe uma verticalização das predições desses dois cenários de degradação, havendo assim, uma incapacidade do modelo de corretamente distinguir entre gradações de congelamento de quadros quando o vídeo com o H.265. Observa-se que o cenário HRC6 possui uma das taxas de bits mais baixas entre todos os cenários de degradação e, por consequência, a compressão mais severa nos vídeos dentre todos os cenários. Concomitantemente à compressão, este cenário possui o maior número de eventos com as maiores durações. Mesmo com degradações mais severas, os vídeos do cenário HRC6 foram preditos com valores de qualidade similares ao HRC8. Além do agrupamento incorreto dos cenários HRC6 e HRC8, os vídeos do cenário ANC2, cenário com a menor taxa de compressão, possui muitas predições com valores inferiores aos valores de MOS. Dentre os cenários com menor degradação, o cenário ANC2 é o caso em que o cenário onde os vídeos estão concentrados em uma região mais distante da diagonal principal.

Na Figura 4.7 (b), é possível notar que os cenários de degradações de perda de pacotes foram melhor divididos do que os cenários de congelamento de quadros para a codificação H.265. Pode-se notar que existe uma diferença mais clara entre as regiões ocupadas por cada cenário contendo perda de pacotes. Além disso, os conjuntos HRC5 e HRC4 encontram-se majoritariamente na região da diagonal principal, indicando que as predições estão mais de acordo com a MOS. No entanto, no cenário HRC2, foram atribuídos valores de qualidade similares para vídeos com valores de MOS diferentes. Existe uma verticalização da dispersão dos vídeos de forma similar a outros cenários de degradações em outros diagramas de dispersão.

Olhando para os quatro diagramas de dispersão, pode-se observar que a NAVEv2 tem maior facilidade em fazer uma distinção entre casos com e sem degradação do que uma distinção entre gradações das degradações. Este comportamento está presente nos gráficos de todas as combinações de compressão e degradação temporal. Outra observação pertinente é que o modelo tem maior dificuldade de detectar uma determinada degradação temporal dependendo do tipo de algoritmo de compressão utilizado. Fica a dúvida se o modelo é capaz de distinguir degradações temporais em cenários equivalentes de compressão.

A Figura 4.8(a) apresenta o diagrama de dispersão para os cenários de degradação HRC8 e HRC9 e os respectivos centroides de cada distribuição. O cenário HRC8 possui taxa de bit igual a 1000 kbits/s, com dois eventos de congelamento de mesma duração e posicionados no meio e no final do vídeo. O cenário HRC9 possui taxa de bit igual a 2000 kbits/, com dois eventos de congelamento posicionados no início e no final do vídeo. Os dois cenários possuem taxas de bit equivalentes e, por este motivo, o efeito da compressão pode ser desprezado e o congelamento de quadros pode ser analisado individualmente. Observa-se que existe uma diferença entre as duas distribuições. As notas de qualidade para os vídeos do cenário HRC8 são inferiores às notas do cenário HRC9. Ambos os cenários possuem quantidade de eventos de congelamento iguais, no entanto, o cenário HRC8 possui os dois eventos mais próximos do final do vídeo, que tendem a ser julgados menos favoravelmente por observadores. Logo a diferença de desempenho entre os dois cenários pode ser atribuída à métrica estar avaliando corretamente a qualidade dos vídeos.

A Figura 4.8(b) apresenta o diagrama de dispersão para os cenários HRC3 e HRC4 e os seus



(a) Congelamento de quadros

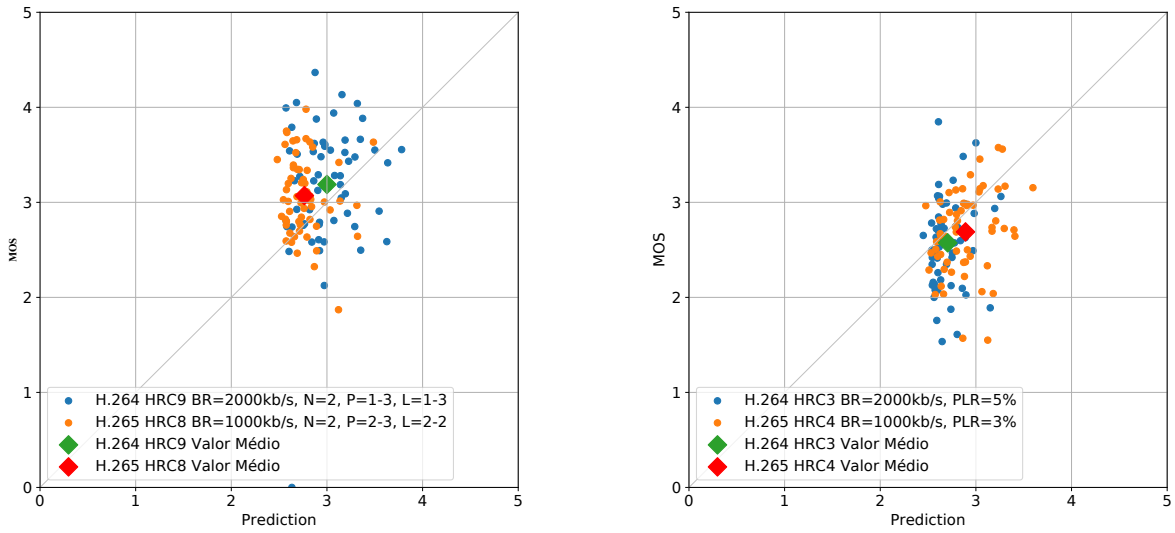
(b) Perda de Pacotes

Figura 4.7: Diagrama de dispersão do vídeos da base Experimento1 codificados em h264 e contendo congelamento de quadros e vídeos codificados em h264 e contendo perda de pacotes.

respectivos centroides. O cenário HRC3 é composto de compressão H.264 com taxa de bits de 2000 kbits/s, com perda de pacotes a uma taxa de 5%. O cenário HRC4 possui compressão H.265 a 1000 kbits/s, com perda de pacotes à taxa de 3%. Na figura, é possível notar que existe uma grande sobreposição entre as distribuições dos dois cenários. Esta sobreposição é explicada pelo fato das degradações dos dois casos serem muito parecidas. Primeiramente, os efeitos de compressão são iguais para os 2 cenários. Além disso, as taxas de perda de pacotes são muito próximas, uma diferença de apenas 2%. Apesar da proximidade entre os dois cenários, existe uma diferença entre os centroides das nuvens de pontos. O cenário HRC3 contém degradações mais intensas e está com um centroide posicionado abaixo do centroide do cenário HRC4. Este resultado é mais um indício do bom funcionamento da NAVEv2.

4.2.4 Inspeção dos Atributos Temporais

Apesar das análises já realizadas nesta seção que apontam que os atributos SinnoM causarem uma melhora do desempenho da NAVE, não se pode visualizar o seu comportamento na presença de degradações. As Figuras 4.9 e 4.10 mostram as curvas dos valores dos atributos pelo tempo para dois cenários de degradação do vídeo v11. Somente o vídeo v11 é analisado para simplificação da análise. Como o número de atributos é grande, a análise é feita em apenas 2 atributos de um total de 8. Como o que diferencia os 8 atributos são as direções das subtrações entre quadros, como explicado na Seção 3.2.2, pode se fazer a análise de apenas 2 atributos de uma direção sem que haja grandes perdas de generalização. Além disso, qualquer diferença causada pela seleção de qualquer uma das direções, é muito influenciada pelo conteúdo de cada sequência. Os atributos



(a) Congelamento de quadros

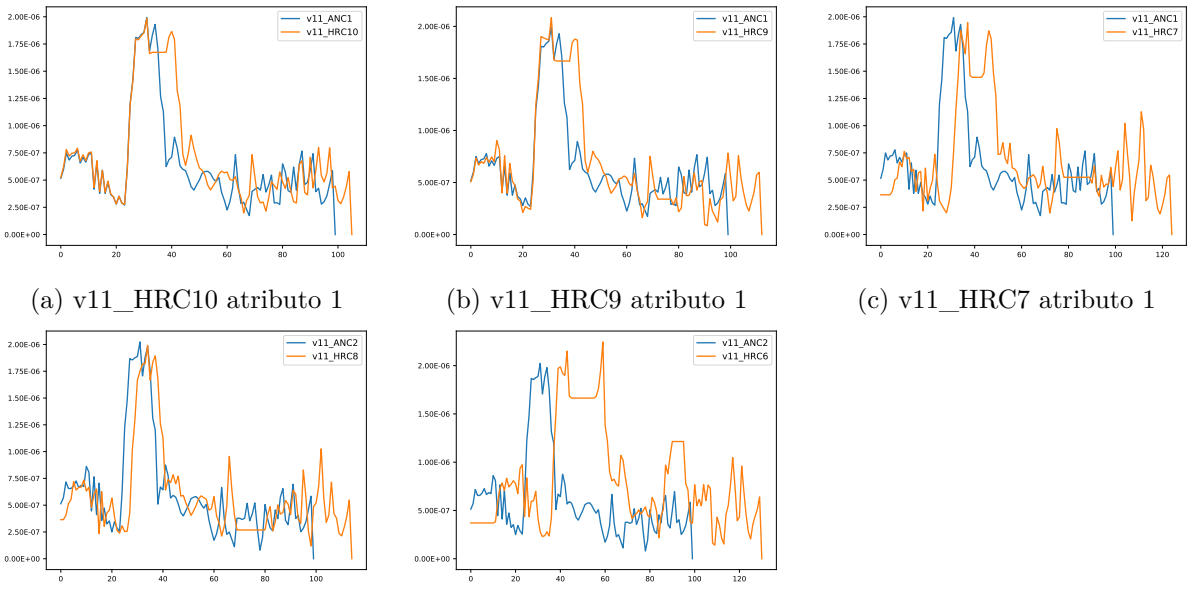
(b) Perda de Pacotes

Figura 4.8: Diagrama de dispersão de dois cenários contendo congelamento de quadros e perda de pacotes equivalentes sendo um codificado em h264 e o outro codificado em h265

escolhidos para a análise são os atributos da direção horizontal. Em cada uma dessas plotagem, o eixo x representa cada os quadros de cada vídeo e o eixo y representa os valores dos atributos temporais.

A Figura 4.9 apresenta o comparativo entre os casos sem degradação e um caso degradado. As subfiguras foram divididas de forma que a primeira linha contém os casos codificados com H.264 e a segunda linha contém os casos codificados com H.265. Nas cinco figuras, é possível notar regiões da curva nos quais o valor dos atributos é constante e se diferencia da curva sem degradação que possui variações de valor. Durante esses intervalos de valor constante ocorrem os eventos de congelamento de quadros. Além disso, é possível notar que os cenários degradados possuem um número maior de quadros, justamente por causa da implementação dessas degradações na base de dados que prolonga o vídeo quando existe congelamento. Em especial, para a sequência de vídeos v11, é possível notar que existe um pico dos valores nesse conteúdo, que é atrasado em decorrência da existência ou não de uma degradação de congelamento no início do vídeo.

Nos cenários onde existe perda de pacotes, é possível notar que as curvas do cenário original e do cenário degradado possuem grande paridade em todos os cinco casos. Nos cenários com perdas de pacote, não existe defasagem entre as duas curvas. As curvas se acompanham e apresentam formatos semelhantes. A observação do instante das degradações temporais se torna menos óbvio. É possível notar que existem variações e oscilações pontuais dos valores do atributo 1 em algumas regiões desses vídeos. Estas variações são facilmente vistas nos diferentes picos presentes, em especial nas Figuras 4.10 (b) e 4.10 (c). Para verificar se estes picos pontuais são eventos de degradações de perda de pacotes, inspecionamos visualmente os quadros correspondentes aos picos onde os cenários degradados possuem valores diferentes da curva do vídeo ANC2. Os quadros



(a) v11_HRC10 atributo 1 (b) v11_HRC9 atributo 1 (c) v11_HRC7 atributo 1

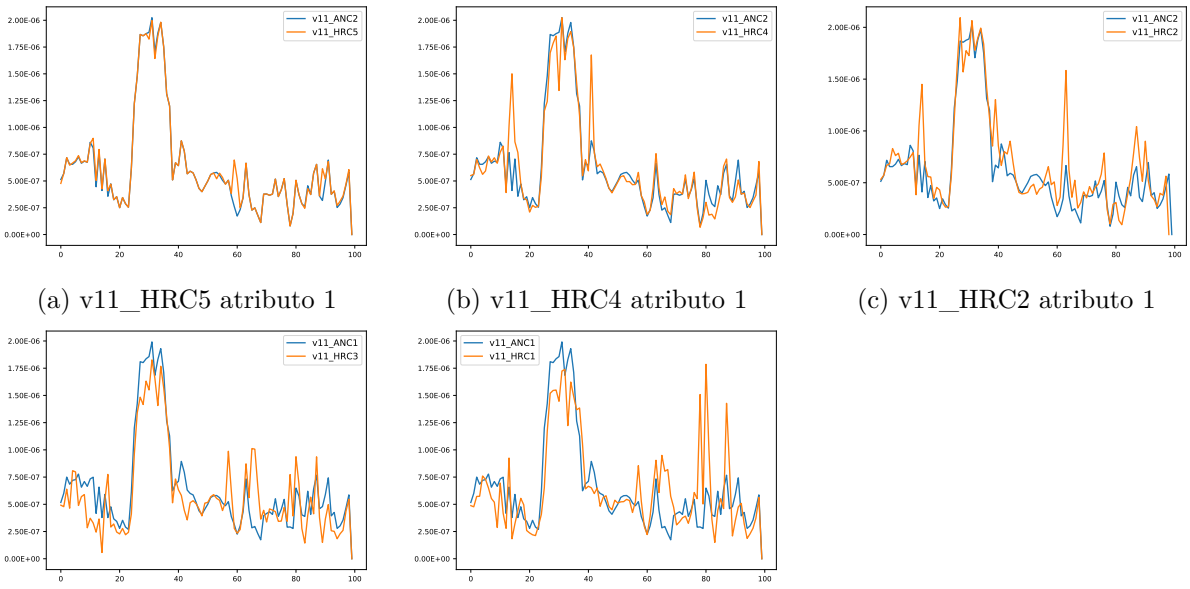
(d) v11_HRC8 atributo 1 (e) v11_HRC6 atributo 1

Figura 4.9: Curvas dos valores dos atributos 1 para os 5 cenários de degradação de congelamento de quadros onde ANC1 e ANC2 são os casos sem degradação codificados respectivamente em H.264 e H.265. As figuras (a), (b) e (c) apresentam curvas codificadas em H.264 e as figuras (d), e (e) apresentam curvas codificadas em H.265.

destes vídeos podem ser vistos na Figura 4.11. A 4.11 (b) apresenta o quadro 425 do vídeo v11_HRC4 que tem fortes distorções de compressão e perda de pacotes. A Figura 4.11 (b) apresenta o quadro 641 do vídeo v11_HRC2 que também possui fortes níveis de degradação.

Nas Figuras 4.10 (d) e 4.10 (e), os picos nas curvas dos vídeos degradados não são tão discrepantes da curva do vídeo ANC1, apesar de existirem oscilações dos valores do atributo da curva do cenário degradado da Figura 4.10 (e). Como as oscilações estão presentes ao longo de todo o vídeo, acredita-se que esta variação seja efeito da compressão. Os quadros destes dois cenários podem ser vistos na Figura 4.11 (c) e (d). O quadro 160 do vídeo v11_HRC3corresponde ao ponto mínimo no começo da curva do atributo 1, enquanto que o quadro 810 do vídeo v11_HRC1 corresponde ao ponto de máximo no final da curva. Pode-se notar que ambos os quadros apresentam fortes degradações. O fato de existirem picos mais claros nas Figuras 4.10 (b) e 4.10 (c) do que nas Figuras 4.10 (d) e 4.10 (e), pode explicar as observações feitas nas análises das Figuras 4.6 e 4.7, onde a percebemos que a NAVEv2 é capaz de distinguir os cenários HRC2 e HRC4 mais facilmente do que os cenários HRC1 e HRC3.

Nas Figuras 4.12 e 4.13, estão apresentadas as curvas dos valores do atributo 2 para os diferentes cenários com e sem degradação. Na Figura 4.12, temos os cenários contendo congelamento de quadros e na Figura 4.13, temos os cenários contendo perdas de pacotes. Na Figura 4.12, podemos observar que existe uma defasagem entre a curva do vídeo base e a curva do vídeo degradado, de forma similar ao ocorrido na Figura 4.9. Novamente, a defasagem ocorre devido aos eventos de congelamento de quadros que atrasam o *display* do conteúdo. De forma semelhante ao ocorrido



(a) v11_HRC5 atributo 1 (b) v11_HRC4 atributo 1 (c) v11_HRC2 atributo 1

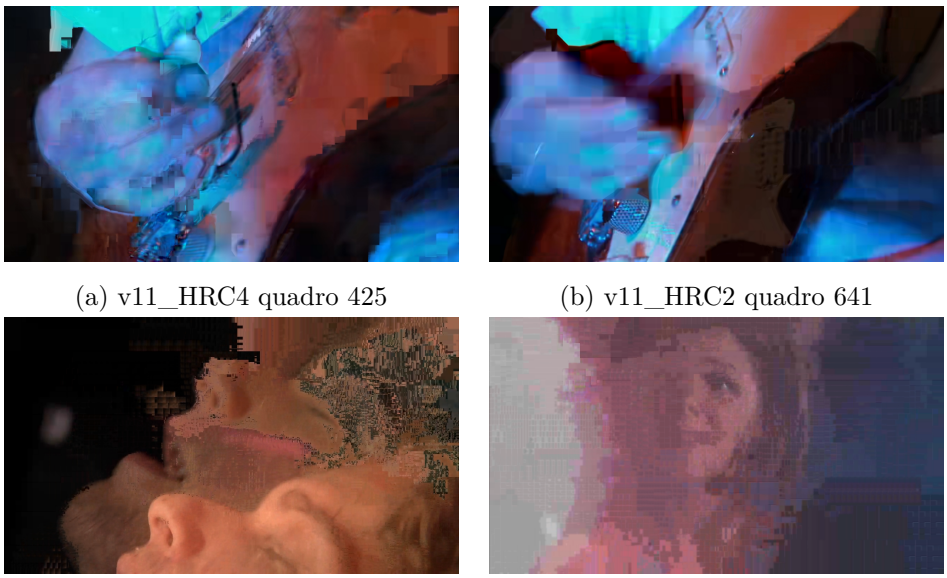
(d) v11_HRC3 atributo 1 (e) v11_HRC1 atributo 1

Figura 4.10: Curvas dos valores dos atributos 1 para os 5 cenários de degradação de perda de pacotes onde ANC1 e ANC2 são os casos sem degradação codificados respectivamente em H.264 e H.265. As figuras (a), (b) e (c) apresentam curvas codificadas em H.265 e as figuras (d), e (e) apresentam curvas codificadas em H.264.

na Figura 4.9, pode-se notar que existem regiões nas quais o valor dos atributos é constante que são os períodos em que ocorrem os eventos de congelamento de quadros.

Outra observação possível é que em alguns cenários os coeficientes do video degradado possuem valores inferiores aos apresentados pelo vídeo original. Por exemplo, no começo do vídeo HRC9, os valores para os primeiros 20 quadros do vídeo possuem valores inferiores aos valores dos primeiros 20 quadros do vídeo base. Este comportamento está presente também para a versão HRC7 e HRC9. Este efeito, no entanto, não está presente nos nas versões HRC10 e HRC8 que são os 2 cenários ue possuem a menor taxa de compressão. A diminuição dos valores de pico dos coeficientes pode ser um efeito da compressão do vídeo. Observe que os valores de pico são reduzidos em todas as regiões dos vídeos, não só em uma porção do vídeo. Logo existe um efeito contínuo ao longo do vídeo, o que não seria o caso para os eventos de degradação temporal.

Na Figura 4.13 são apresentados os gráficos contendo as degradações de perdas de pacotes. É possível notar que, de forma similar a Figura 4.10, não existe um atraso entre as duas curvas, onde a primeira é a curva do vídeo âncora, destacada em azul, e a segunda é a curva do vídeo degradado. Existe uma semelhança muito grande entre as curvas do vídeo degradado e do vídeo original, em especial para o cenário HRC5. No entanto, para as outras curvas, pode-se notar que existe uma maior diferenciação entre a versão degradada e o conteúdo original. Nas Figuras 4.13 (d) e (e), que contém respectivamente os sinais HRC3 e HRC1, nota-se que existe uma maior diferença entre os valores máximos das duas curvas. Além da redução dos valores máximos, existe uma translação da curva para valores inferiores nos 2 cenários degradados. Através dos quadros



(a) v11_HRC4 quadro 425 (b) v11_HRC2 quadro 641

(c) v11_HRC3 quadro 160 (d) v11_HRC1 quadro 810

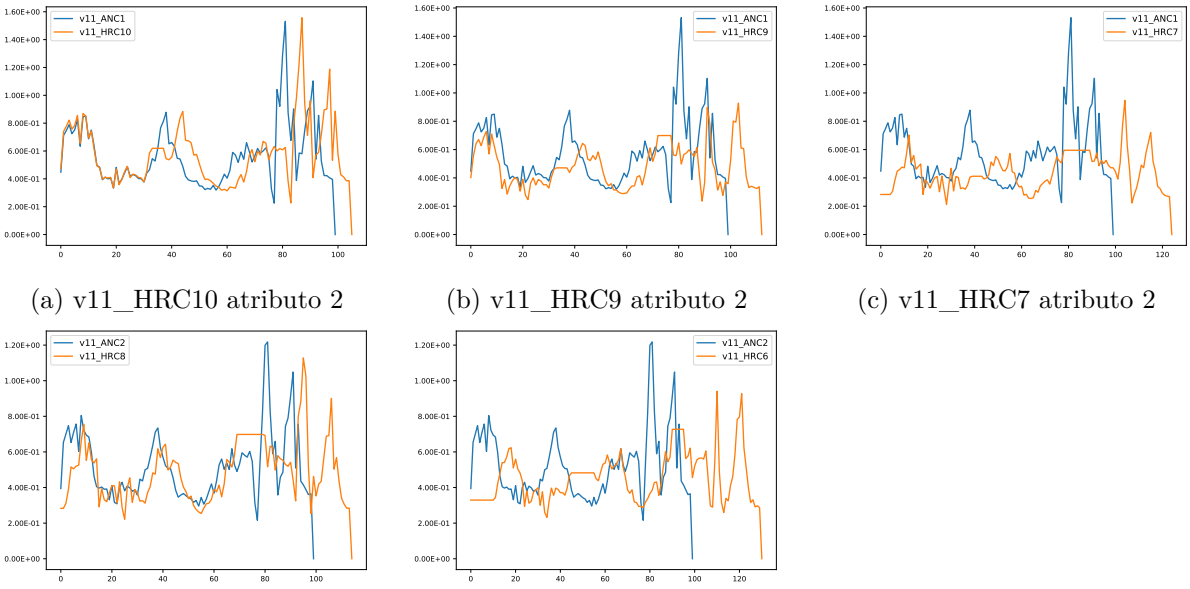
Figura 4.11: Quadros dos cenários degradados das referentes aos pontos onde a curva do cenários degradados divergem da curva do cenário sem degradações. As figuras (a) e (b) apresentam quadros codificadas em H.265 e as figuras (c) e (d) apresentam quadros codificadas em H.264.

apresentados nas Figuras 4.14 (c) e (d), observa-se a existência de um alto grau de degradação nestes instantes dos vídeos. No entanto, o atributo 2 de cada um dos quadros não acompanha o nível de degradação, sendo um indício para a baixa sensibilidade do atributo 2 às degradações temporais.

Como dito anteriormente, não existem grandes variações entre a curva do vídeo original e as curvas dos vídeos dos cenários HRC5, HRC4 e HRC2. Nota-se que as variações ocorridas no atributo 2 para os vídeos codificados com H.265 são referentes à variações do conteúdo, não tendo relação com o seu nível de degradação. Isto se deve à semelhança da curva do cenário degradado com a curva do cenário não degradado. Nos cenários codificados com H.264, observa-se que as curvas para os cenários degradados possuem poucas variações ao longo do tempo, se comparada com o cenário original. Desta forma, pode-se concluir que o atributo 2 é incapaz de diferenciar entre os instantes do vídeo onde as degradações acontecem, atribuindo valores semelhantes para quadros independentemente da sua qualidade.

4.3 Testes na base LIVE-NETFLIX-II

Diferentemente do experimento um realizado na base de dados da UNB, os testes realizados na base de dados LIVE-NETFLIX-II foram realizados apenas para dois conjuntos de atributos. Dentre os conjuntos de atributos já testados, escolhemos o conjunto 1, que contém apenas os atributos espaciais DIIVINE, e o conjunto 4, que contém os atributos DIIVINE e SinnoM. O conjunto 1 foi escolhido por se assemelhar ao conjunto de atributos originais da NAVE. No entanto, sabe-se que o desempenho da NAVE com apenas os atributos espaciais é levemente superior ao



(a) v11_HRC10 atributo 2 (b) v11_HRC9 atributo 2 (c) v11_HRC7 atributo 2

(d) v11_HRC8 atributo 2 (e) v11_HRC6 atributo 2

Figura 4.12: Curvas dos valores dos atributos 2 para os 5 cenários de degradação de congelamento de quadros onde ANC1 e ANC2 são os casos sem degradação codificados respectivamente em H.264 e H.265. As figuras (a), (b) e (c) apresentam curvas codificadas em H.264 e as figuras (d), e (e) apresentam curvas codificadas em H.265.

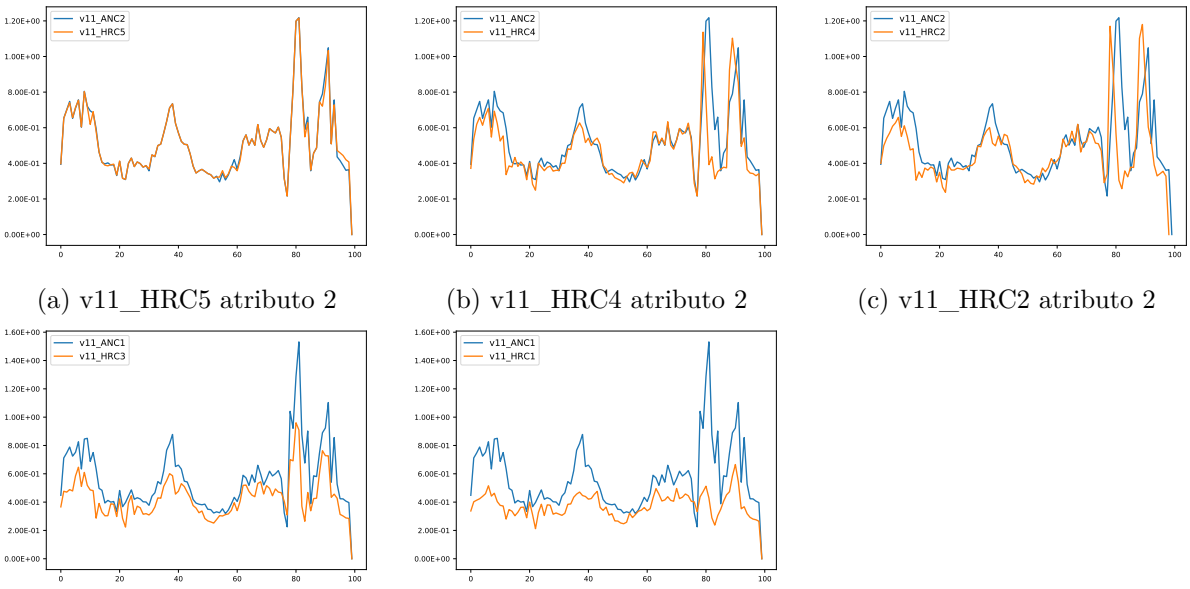
	Diivine	**Diivine+SinnoM**
PCC	0.931	0.936
SCC	0.918	0.918
RMSE	0.621	0.604

Tabela 4.4: Resultados dos valores de PCC, o SCC e o RMSE para dois conjuntos de atributos, testados na base de dados LIVE-Netflix-II.

desempenho do método original. Essa diferença, entretanto, não invalida a comparação com o novo conjunto de atributos. O segundo conjunto de atributos é composto pelo conjunto de atributos selecionados durante os experimentos com a base de dados anterior para compor a NAVEv2. Estes testes com a segunda base de dados foram realizados com o intuito de validar o conjunto de atributos selecionados para a NAVEv2. Mas, é preciso enfatizar que as duas bases de dados não possuem os mesmos tipos de degradações.

Nos testes realizados na base de dados LIVE-Netflix-II, a base de dados foi separada em 5 partições e o treinamento e o teste ocorreu utilizando a técnica de *k-fold*. A Tabela 4.4 contém a média do PCC, SCC e RMSE para os testes.

As Figuras 4.15 (a), (b) e (c) apresentam, respectivamente, o PCC, o SCC e o RMSE para as 5 partições de teste. Na Figura 4.15 (a), observamos que o PCC de ambos os conjuntos têm valores altos de correlação. O conjunto contendo os atributos DIIVINE e SinnoM obteve uma distribuição mais compacta e com valores médios superiores ao conjunto que é composto apenas



(a) v11_HRC5 atributo 2 (b) v11_HRC4 atributo 2 (c) v11_HRC2 atributo 2

(d) v11_HRC3 atributo 2 (e) v11_HRC1 atributo 2

Figura 4.13: Curvas dos valores dos atributos 2 para os 5 cenários de degradação de congelamento de quadros onde ANC1 e ANC2 são os casos sem degradação codificados respectivamente em H.264 e H.265. As figuras (a), (b) e (c) apresentam curvas codificadas em H.265 e as figuras (d), e (e) apresentam curvas codificadas em H.264.

pelos atributos DIIVINE. Além disso, sua distribuição mais compacta com uma mediana de valor mais alto, sendo equivalente ao terceiro quartil da distribuição dos atributos DIIVINE. Contudo, os valores superiores do conjunto 4 não conseguiram superar os valores superiores do conjunto 1.

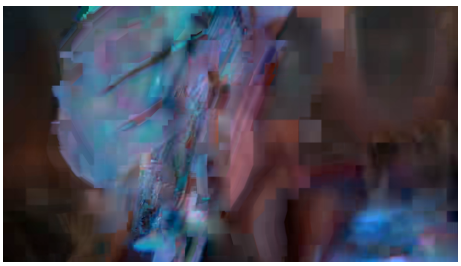
É possível notar na Figura 4.15 (b) que ambos os conjuntos novamente obtiveram altos valores de correlação com a MOS. O conjunto 1 obteve uma distribuição mais dispersa do que o conjunto 4. Outro ponto válido de se ressaltar é que a distribuição do conjunto 4 está concentrada em valores superiores da distribuição, onde 50% da distribuição do conjunto 4 encontra-se acima da mediana do conjunto 1. Na Figura 4.15 (c), a distribuição RMSE para o conjunto 4 possui valores inferiores aos valores da distribuição para a métrica treinada com o conjunto 1 de atributos.

4.4 Discussões Finais

Neste capítulo, fizemos uma análise de diversos conjuntos de atributos candidatos para substituir os atributos temporais utilizados na arquitetura original da NAVE. Analisamos os diferentes conjuntos de atributos candidatos quanto ao valor médio e quanto a distribuição dos coeficientes de correlação. A NAVEv2 foi comparada à diversas métricas com e sem referência dentre elas a NAVE e obteve desempenho superior a todas as métricas.

Em seguida, analisamos a os diagramas de caixa com as distribuições dos coeficientes de correlação. Durante estas análises observamos que a inserção dos novos atributos temporais não causou grandes aumentos do valor médio da correlação. No entanto, a distribuição dos valores

(b) v11_HRC4 quadro 550



(c) v11_HRC2 quadro 395

(e) v11_HRC3 quadro 365

(e) v11_HRC1 quadro 800

Figura 4.14: Quadros dos cenários degradados das referentes aos pontos onde a curva do cenários degradados divergem da curva do cenário sem degradações. As figuras (a) e (b) apresentam quadros codificadas em H.265 e as figuras (c) e (d) apresentam quadros codificadas em H.264.

das correlações apresenta um aumento dos valores mínimos obtidos pela métrica em relação à arquitetura original.

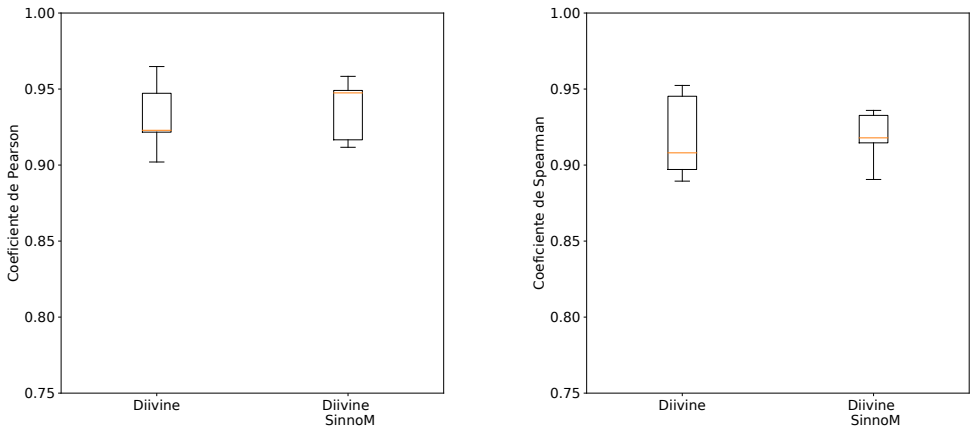
Devido à limitações da base de dados Qualidade Audiovisual UnB 2018 (Experimento 1), não foi possível verificar o desempenho da métrica em cenários onde as degradações temporais estão isoladas da compressão. Todos os cenários são compostos de uma degradação temporal e um algum tipo de compressão. Por este motivo, fizemos a análise dos diagramas de dispersão para os diferentes cenários de degradação. Separando os diferentes tipos de compressão e os tipos de degradações temporais, observamos que a qualidade predita para vídeos com determinados tipos de degradação temporal varia de acordo com o tipo de algoritmo de compressão utilizado. Observou-se que a NAVEv2 conseguiu, em grande parte, distinguir entre os diversos casos. No entanto, não foi possível observar o comportamento da métrica na presença de degradações temporais isoladamente.

Como não havia a disponibilidade de cenários de degradação contendo somente as degradações temporais, como o congelamento de quadros e perda de pacotes, usou-se cenários da base de dados cujas taxas de bit da compressão eram equivalentes. Realizamos uma análise comparativa destes dois cenários para entender o comportamento da métrica quando apenas a degradação temporal era alterada. Nas comparações, notamos que apesar da baixa diferença de degradação temporal, a NAVEv2 foi capaz de atribuir estimativas de qualidade condizentes com o nível de degradação de cada cenário.

Também analisamos o comportamento dos 2 atributos temporais ao longo do tempo. Fez-se a plotagem das curvas dos valores dos atributos ao longo do tempo. Analisando as curvas dos valores dos descritores, identificamos onde acontecem os eventos de congelamento de quadros.

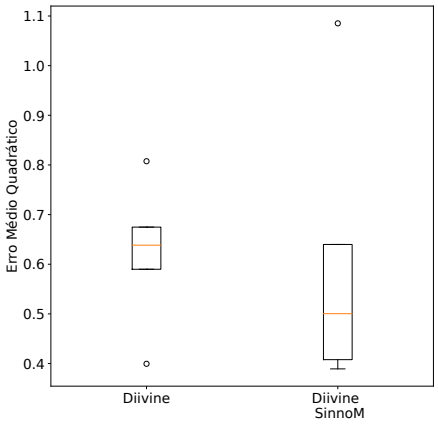
Por outro lado, não foi possível identificar tão facilmente os pontos onde acontecem os eventos de perda de pacotes, apesar de alguns picos nas curvas puderem ser atribuídos a degradações. Observamos também que existem comportamentos diferentes nas curvas nos cenários degradados de cada atributo dependendo do tipo de codificador que é utilizado.

Finalmente, também treinamos e testamos a métrica NAVEv2 na base de dados da LIVE-Netflix-II. Foi possível notar que as predições da métrica possuem alta correlação com os escores de qualidade. Além disso, a métrica NAVEv2 obteve um resultado ligeiramente melhor do que a arquitetura da NAVE, que utiliza apenas atributos espaciais.



(a) PCC

(b) SCC



(c) RMSE

Figura 4.15: Diagrama de caixa do PCC, SCC e RMSE para todas as degradações da Live-Netflix-II para 2 conjunto de atributos.

Capítulo 5

Conclusões

Este capítulo apresenta as conclusões feitas a partir dos dados e análises apresentadas nos capítulos anteriores. Além das conclusões também são dispostas propostas de trabalhos futuros que possam se tornar próximas contribuições para esta área de pesquisa.

5.1 Contribuições

Neste trabalho, estudamos o aspecto temporal do desempenho da métrica de qualidade NAVE, buscando expandir o número de atributos temporais relevantes e descartar atributos que não contribuíam para o desempenho da métrica. Estudamos o comportamento da métrica e dos atributos temporais em diferentes cenários de degradação. Para isto, fizemos diversos estudos utilizando a base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1).

No primeiro estudo, exploramos as degradações em conjunto. Analisamos o desempenho da métrica para diferentes conjuntos de atributos e exploramos o uso de novos atributos (isoladamente ou em combinação com outros tributos). Observamos que os atributos BRISQUE não foram capazes de capturar informações sobre as degradações temporais, ao contrário dos atributos Sinno. Dentre as duas formas de processamento dos atributos Sinno, percebemos que os atributos SinnoM geraram melhores resultados tanto para a base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1) como para a base de dados LIVE-Netflix-II. Além disso, comparamos o desempenho da métrica com os melhores atributos (NAVEv2) com outras métricas com e sem referência.

O segundo estudo foi feito com as degradações temporais isoladas. Separamos os cenários da base de dados entre os diferentes algoritmos de compressão e tipos de degradação temporal. Observamos que NAVEv2 foi capaz de atribuir estimativas de qualidade melhores para alguns cenários do que outros. Percebe-se que as combinações entre o tipo de codificador e o tipo de degradação temporal afetam o desempenho da métrica. Analisando as degradações temporais isoladamente, comparamos os cenários onde os codificadores H.264 e H.265 estavam com taxas de bit equivalentes, ou seja, onde não haveria diferença entre as degradações de compressão. Nestes cenários, pode-se notar que a métrica atribuiu (em média) notas de qualidade diferentes para os dois cenários.

Através das curvas dos valores dos atributos 1 e 2, é possível entender melhor o comportamento dos atributos frente a degradações. Observamos que ambos os atributos são capazes de detectar as degradações de congelamento de quadros. Nestes cenários, os valores permanecem constantes durante o evento de congelamento. Para os cenários de perda de pacotes, observamos que um dos atributos é sensível a degradação. Nos instantes de degradação, os valores do atributo variam abruptamente. No entanto, o outro tipo de atributo se mostrou insensível foi incapaz de diferenciar cenários degradados de cenários não degradados. Através dos resultados dos testes feitos nas duas bases de dados, foi possível coletar diferentes resultados, permitindo afirmar que os novos atributos temporais escolhidos possuem um melhor desempenho na avaliação da qualidade de vídeo com degradações temporais. No entanto, mostrou-se necessário a adição de uma etapa de pós processamento dos atributos para aumentar a capacidade de discriminação entre cenários degradados e cenários não degradados.

5.2 Trabalhos Futuros

Para trabalhos futuros, pode-se testar novos atributos temporais. No decorrer deste trabalho, foram utilizadas apenas 2 versões dos atributos Sinno. Além disso, todas os atributos testados usam a abordagem NSS. Atributos baseados em outras abordagens podem trazer bons resultados na métricas. Um segundo tópico para trabalhos futuros é o pós-processamento dos atributos temporais. Como foi possível observar nos experimentos com os diferentes conjuntos de atributos, houve uma diferença de desempenho da métrica para os conjuntos que utilizam os atributos SinnoM e SinnoP. Os atributos SinnoP são compostos pelos coeficientes dos diferentes macro blocos de cada quadro, enquanto que os atributos SinnoM correspondem à média dos valores dos macro-blocos do quadro. Esse simples pós processamento dos atributos causou uma melhora do desempenho. Além disso, as curvas dos atributos temporais SinnoM para o vídeo v11 mostraram que as degradações de congelamento de quadros causam valores constantes nos atributos e degradações de perdas de pacotes causam picos de intensidade. Sendo assim, seria possível realizar um pós-processamento desses atributos para que as regiões constantes fossem evidenciadas de forma a capturar melhor os eventos de congelamento de quadros. De forma complementar, poderia-se gerar outro conjunto de atributos pós-processados que evidenciasse os picos de forma a melhorar a detecção de eventos de perda de pacotes.

Outra área para trabalhos futuros que pode ser investigada é o desempenho positivo dos atributos espaciais na avaliação da qualidade dos vídeos contendo degradações temporais. A hipótese para o bom desempenho da NAVE com somente os atributos espaciais consiste no fato da base de dados de Qualidade Audiovisual UnB 2018 (Experimento 1) conter apenas cenários com conjuntos de degradações. Os cenários são compostos por conjuntos de degradação de compressão degradações de perda de pacotes ou congelamento de quadros. Apesar dos testes realizados, pode ser que os atributos espaciais estejam na verdade identificando as degradações de compressão e não necessariamente as degradações temporais. É válido ressaltar que em todos os cenários da base de dados, ocorrem acréscimos simultâneos de compressão e de degradações temporal. Desta forma, é difícil avaliar degradações temporais isoladamente, sendo necessária a utilização de outras

bases de dados especializadas em degradações temporais isoladas.

Outra linha de pesquisa possível é a otimização da arquitetura da métrica nave. Neste trabalho, não foi possível propor alterações na arquitetura da métrica. Logo, a arquitetura e o treinamento da rede permaneceram iguais aos da métrica original. Seria possível melhorar o desempenho da métrica com um melhoramento do processo de treinamento e um estudo mais aprofundado do ajuste dos seus hiper parâmetros.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] GONZALEZ, R. C. *Processamento Digital de Imagens*. 3rd. ed. [S.l.]: Pearson, 2009.

[2] MARTINEZ, H. B.; FARIAS, M. C.; HINES, A. A no-reference autoencoder video quality metric. In: IEEE. *2019 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2019. p. 1755–1759.

[3] PIAMRAT, K. et al. Quality of experience measurements for video streaming over wireless networks. In: IEEE. *2009 Sixth International Conference on Information Technology: New Generations*. [S.l.], 2009. p. 1184–1189.

[4] MARTINEZ, H. A. B. A Three Layer System for Audio-visual Quality Assessment. 2019.

[5] ITU-T. H.264 : Advanced video coding for generic audiovisual services. *Technical report*, 2003.

[6] ITU-T. H.265 : High efficiency video coding. *Technical report*, 2013.

[7] UNTERWEGER, A. Compression artifacts in modern video coding and state-of-the-art means of compensation. In: *Multimedia Networking and Coding*. [S.l.]: IGI Global, 2013. p. 28–49.

[8] GONZALEZ, R. E. W. R. C.; HALL, P. *Digital Image Processing (2nd Edition)*. [S.l.]: Tom Robbins.

[9] BAMPIS, C. G.; BOVIK, A. C. Learning to predict streaming video qoe: Distortions, rebuffering and memory. *arXiv preprint arXiv:1703.00633*, 2017.

[10] SESHADRINATHAN, K. et al. Study of subjective and objective quality assessment of video. *IEEE transactions on Image Processing*, IEEE, v. 19, n. 6, p. 1427–1441, 2010.

[11] WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, IEEE, v. 26, n. 1, p. 98–117, 2009.

[12] WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.

[13] AABED, M. A.; ALREGIB, G. Peqaso: Perceptual quality assessment of streamed videos using optical flow features. *IEEE Transactions on Broadcasting*, IEEE, v. 65, n. 3, p. 534–545, 2018.

[14] AKHTAR, Z.; FALK, T. H. Audio-visual multimedia quality assessment: A comprehensive survey. *IEEE access*, IEEE, v. 5, p. 21090–21117, 2017.

[15] YANG, K.-C. et al. Perceptual temporal quality metric for compressed video. *IEEE Transactions on Multimedia*, IEEE, v. 9, n. 7, p. 1528–1535, 2007.

[16] MITTAL, A.; SAAD, M. A.; BOVIK, A. C. A completely blind video integrity oracle. *IEEE Transactions on Image Processing*, IEEE, v. 25, n. 1, p. 289–300, 2015.

[17] RUDERMAN, D. L. The statistics of natural images. *Network: computation in neural systems*, IOP Publishing, v. 5, n. 4, p. 517, 1994.

[18] KEIMEL, C. et al. Video is a cube. *IEEE Signal Processing Magazine*, IEEE, v. 28, n. 6, p. 41–49, 2011.

[19] YANG, J. et al. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 26, n. 1, p. 131–137, 2004.

[20] SINNO, Z.; BOVIK, A. C. Spatio-temporal measures of naturalness. In: IEEE. *2019 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2019. p. 1750–1754.

[21] HORN, B. K.; SCHUNCK, B. G. "determining optical flow": A retrospective. Elsevier, 1993.

[22] KORHONEN, J. Two-level approach for no-reference consumer video quality assessment. *IEEE Transactions on Image Processing*, IEEE, v. 28, n. 12, p. 5923–5938, 2019.

[23] VU, P. V.; CHANDLER, D. M. Vis3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices. *Journal of Electronic Imaging*, International Society for Optics and Photonics, v. 23, n. 1, p. 013016, 2014.

[24] LARSON, E. C.; CHANDLER, D. M. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, International Society for Optics and Photonics, v. 19, n. 1, p. 011006, 2010.

[25] LUCAS, B. D.; KANADE, T. et al. An iterative image registration technique with an application to stereo vision. In: VANCOUVER, BRITISH COLUMBIA. [S.l.], 1981.

[26] ZHANG, Y. et al. C-diivine: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes. *Signal Processing: Image Communication*, Elsevier, v. 29, n. 7, p. 725–747, 2014.

[27] OSTASZEWSKA, A.; KŁODA, R. Quantifying the amount of spatial and temporal information in video test sequences. In: *Recent Advances in Mechatronics*. [S.l.]: Springer, 2007. p. 11–15.

[28] MOORTHY, A. K.; BOVIK, A. C. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, IEEE, v. 20, n. 12, p. 3350–3364, 2011.

[29] SIMONCELLI, E. P. et al. Shiftable multiscale transforms. *IEEE transactions on Information Theory*, IEEE, v. 38, n. 2, p. 587–607, 1992.

[30] RAO, R. P.; OLSHAUSEN, B. A.; LEWICKI, M. S. *Probabilistic models of the brain: Perception and neural function*. [S.l.]: MIT press, 2002.

[31] SINNO, Z.; BOVIK, A. C. Large-scale study of perceptual video quality. *IEEE Transactions on Image Processing*, IEEE, v. 28, n. 2, p. 612–627, 2018.

[32] MITTAL, A.; SOUNDARARAJAN, R.; BOVIK, A. C. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, IEEE, v. 20, n. 3, p. 209–212, 2012.

[33] LASMAR, N.-E.; STITOU, Y.; BERTHOUMIEU, Y. Multiscale skewed heavy tailed model for texture analysis. In: IEEE. *2009 16th IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2009. p. 2281–2284.

[34] MITTAL, A.; MOORTHY, A. K.; BOVIK, A. C. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, IEEE, v. 21, n. 12, p. 4695–4708, 2012.

[35] MARTINEZ, H. B.; FARIAS, M. C. Using the immersive methodology to assess the quality of videos transmitted in udp and tcp-based scenarios. *Electronic Imaging*, Society for Imaging Science and Technology, v. 2018, n. 12, p. 233–1, 2018.

[36] BOYCE, J. M.; GAGLIANELLO, R. D. Packet loss effects on mpeg video sent over the public internet. In: *Proceedings of the sixth ACM international conference on Multimedia*. [S.l.: s.n.], 1998. p. 181–190.

[37] WENGER, S. H. 264/avc over ip. *IEEE transactions on circuits and systems for video technology*, IEEE, v. 13, n. 7, p. 645–656, 2003.

[38] BAMPIS, C. G. et al. Towards perceptually optimized end-to-end adaptive video streaming. *arXiv preprint arXiv:1808.03898*, 2018.

[39] WANG, Z.; SIMONCELLI, E. P.; BOVIK, A. C. Multiscale structural similarity for image quality assessment. In: IEEE. *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. [S.l.], 2003. v. 2, p. 1398–1402.

[40] MOORTHY, A.; BOVIK, A. A modular framework for constructing blind universal quality indices. *IEEE Signal Processing Letters*, v. 17, 2009.