



UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
BACHARELADO EM ENGENHARIA ELÉTRICA - TELECOMUNICAÇÕES

Avaliação da qualidade de experiência em
transmissão de vídeo com codec H.264 utilizando
métricas objetivas de referência completa e sem
referência

Inara Karyna Torres Batista

Manaus – AM
Novembro de 2023

Inara Karyna Torres Batista

Avaliação da qualidade de experiência em transmissão de vídeo com codec H.264 utilizando métricas objetivas de referência completa e sem referência

Monografia de Graduação apresentada a Faculdade de Tecnologia da Universidade Federal do Amazonas como requisito parcial para a obtenção do grau de bacharel em Engenharia Elétrica-Telecomunicações.

Orientador

Prof. Dr. Waldir Sabino da Silva Junior

Universidade Federal do Amazonas – UFAM

Faculdade de Tecnologia – FT

Manaus-AM

Novembro de 2023

Ficha Catalográfica

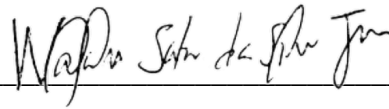
Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

B333a Batista, Inara Karyna Torres
Avaliação da qualidade de experiência em transmissão de vídeo com codec H.264 utilizando métricas objetivas de referência completa e sem referência / Inara Karyna Torres Batista . 2023
56 f.: il. color; 31 cm.

Orientador: Waldir Sabino da Silva Junior
TCC de Graduação (Engenharia Elétrica - Telecomunicações) -
Universidade Federal do Amazonas.

1. Avaliação de experiência. 2. Avaliação de qualidade de vídeo.
3. Transmissão de vídeo. 4. Codec H.264. I. Silva Junior, Waldir Sabino da. II. Universidade Federal do Amazonas III. Título

Monografia de Graduação sob o título Avaliação da qualidade de experiência em transmissão de vídeo com codec H.264 utilizando métricas objetivas de referência completa e sem referência apresentada por Inara Karyna Torres Batista e aceita pela Faculdade de Tecnologia da Universidade Federal do Amazonas, sendo aprovada por todos os membros da banca examinadora abaixo especificada:



Prof. Dr. Waldir Sabino da Silva Junior

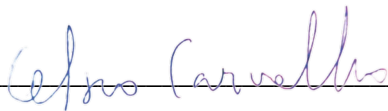
Orientador

Universidade Federal do Amazonas



Prof. MSc. Ademir de Jesus Lourenço

Universidade Federal do Amazonas



Prof. Dr. Celso Barbosa Carvalho

Universidade Federal do Amazonas

Manaus-AM, 06 de novembro de 2023.

Dedico este trabalho aos meu pais

Nildete e Aurivelto

Agradecimentos

Agradeço primeiramente a Deus, por ter me dado força e sabedoria para enfrentar a caminhada até aqui.

Agradeço ao meus pais, Nildete e Aurivelto, por terem me dado o apoio e suporte necessário para alcançar esse tão sonhado objetivo. Agradeço a minha avó Ercília (In Memoriam) por ter sido tão paciente e amorosa quando precisei de seus cuidados.

Agradeço o meu companheiro de vida Patrick, pelas palavras de incentivo, pelo cuidado durante a maior parte do período da faculdade ao qual enfrentamos juntos, e pela companhia nas noites em claro enquanto escrevia esta monografia.

Agradeço também a todos os amigos que fizeram parte dessa trajetória, em especial Elbesson, Julis e Yasmin. Os momentos vividos com vocês pelos corredores, salas de aulas e laboratórios jamais serão esquecidos.

Agradeço ao meu orientador, prof. Dr. Waldir Sabino pelos conselhos e incentivo essenciais para a conclusão deste trabalho. Agradeço também o prof. Dr. Ayres Mardem (In Memoriam) por ter passado seus conhecimentos com tanta dedicação e ter servido de inspiração para muitos de nós.

Avaliação da qualidade de experiência em transmissão de vídeo com codec H.264 utilizando métricas objetivas de referência completa e sem referência

Autor: Inara Karyna Torres Batista

Orientador(a): Prof. Dr. Waldir Sabino da Silva Junior

RESUMO

Com a crescente disponibilização de conteúdo digital nos mais diferentes meios de comunicação, se torna cada vez mais essencial para os provedores garantirem a qualidade do material ofertado, sejam eles textos, imagens ou vídeos. A qualidade de experiência (QoE) realiza essa tarefa, pois permite a avaliação da qualidade percebida pelo usuário. Quando se trata de vídeos, a avaliação de qualidade de vídeo (VQA) permite obter uma estimativa da qualidade perceptiva do telespectador, através das chamadas métricas de avaliação de qualidade. A VQA é amplamente utilizada para obter a qualidade de transmissões de vídeos, permitindo adaptações e melhorias no sistema utilizado. Por sua vez, o codec H.264 é um grande aliado nessas transmissões, permitindo a compactação e maior eficiência dos sistemas utilizados para tal. Este trabalho visa avaliar a qualidade de experiência de vídeos transmitidos com o codec H.264 utilizando dois tipos específicos de métricas, as de referência completa e sem referência.

Palavras-chave: avaliação de experiência, avaliação de qualidade de vídeo, transmissão de vídeo, codec H.264.

Assessment of Quality of Experience in H.264 Video Transmission Using Full-Reference and No-Reference Objective Metrics

Author: Inara Karyna Torres Batista

Advisor: Waldir Sabino da Silva Junior, Ph.D

ABSTRACT

With the increasing availability of digital content across various communication channels, it becomes increasingly essential for providers to ensure the quality of the material they offer, be it text, images, or videos. Quality of Experience (QoE) accomplishes this task by allowing the evaluation of the quality as perceived by the user. In the context of videos, Video Quality Assessment (VQA) provides an estimate of the perceptual quality experienced by viewers through the use of quality assessment metrics. VQA is widely employed to assess the quality of video transmissions, facilitating adaptations and improvements in the systems used. In turn, the H.264 codec plays a significant role in these transmissions, enabling compression and greater efficiency in the systems employed for this purpose. This study aims to evaluate the quality of experience of videos transmitted using the H.264 codec, employing two specific types of metrics: full-reference and no-reference metrics.

Keywords: quality of experience, video quality assessment, video transmission, codec H.264.

Lista de figuras

Figura 1 – Imagem original e imagem com o artefato blurring.	24
Figura 2 – Imagem original e imagem com o artefato blocking.	25
Figura 3 – Processo de codificação	27
Figura 4 – Processo de decodificação	28
Figura 5 – Diagrama do sistema proposto.....	31
Figura 6 –Interface do MSU VQMT	35
Figura 7 – Gráfico de saída produzidos pelo MSU VQMT	36
Figura 8 – Frame 100 do vídeo Shake N Dry original	37
Figura 9 – Frame 100 do vídeo Shake N Dry transmitido.	38
Figura 10 – À esquerda, o quadro 100 do vídeo Shak N Dry transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM	38
Figura 11 – Na imagem acima, o gráfico PSNR para o vídeo Shake N Dry transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Shake N Dry transmitido	39
Figura 12 – À esquerda, o quadro 100 do vídeo transmitida com a métrica Blocking. À direita, o mesmo quadro com a métrica Blurring.....	39
Figura 13 – Na imagem acima, o gráfico Blocking para o vídeo Shake N Dry transmitido. Na imagem abaixo, o gráfico Blurring para o vídeo Shake N Dry transmitido.....	40
Figura 14 – Quadro 200 do vídeo Honey Bee original	41
Figura 15 – Quadro 200 do vídeo Honey Bee transmitido.....	41
Figura 16 – À esquerda, o quadro 200 do vídeo Honet Bee transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM	42
Figura 17 – Na imagem acima, o gráfico PSNR para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Honey Bee transmitido.	42
Figura 18 – Na imagem acima, o gráfico Blocking para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico Blocking para o vídeo Honey Bee transmitido.....	43

Figura 19 – À esquerda, o quadro 200 do vídeo Honey Bee transmitido com a métrica Blocking. À direita, o mesmo quadro com a métrica Blurring	43
Figura 20 – Quadro 200 do vídeo Jockey original.	44
Figura 21 – Quadro 200 do vídeo Jockey transmitido.	44
Figura 22 – À esquerda, o quadro 200 do vídeo Jockey transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM.	45
Figura 23 – Na imagem acima, o gráfico PSNR para o vídeo Jockey transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Jockey transmitido	45
Figura 24 – À esquerda, o quadro 100 do vídeo Jockey transmitido com a métrica Blocking. À direita, o mesmo quadro com a métrica Blurring	46
Figura 25 – Na imagem acima, o gráfico PSNR para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Honey Bee transmitido	46
Figura 26 – Quadro 200 do vídeo Beauty transmitido.	47
Figura 27 – Quadro 200 do vídeo Beauty original.	47
Figura 28 – À esquerda, o quadro 200 do vídeo Beauty transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM	48
Figura 29 – Na imagem acima, o gráfico PSNR para o vídeo Beauty transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Beauty transmitido.	48
Figura 30 – À esquerda, o quadro 100 do vídeo Beauty transmitido com a métrica Blocking. À direita, o mesmo quadro com a métrica Blurring.	49
Figura 31 – Na imagem acima, o gráfico PSNR para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Honey Bee transmitido.	49

Lista de tabelas

Tabela 1 - Classificação ACR	19
Tabela 2 – Classificação DCR	19
Tabela 3 - Valores de referência para a métrica PSNR.	50
Tabela 4 - Valores de referência para a métrica SSIM.....	50
Tabela 5 - Valores de referência para a métrica Blocking.	50
Tabela 6 - Valores de referência para a métrica Blurring.	50
Tabela 7 - Valores obtidos para o vídeo ShakeNDry	51
Tabela 8 - Valores obtidos para o vídeo HoneyBee.....	51
Tabela 9 - Valores obtidos para o vídeo Jockey.....	51
Tabela 10 - Valores obtidos para o vídeo Beauty.	52

Lista de abreviaturas e siglas

UFAM – Universidade Federal do Amazonas

FT – Faculdade de Tecnologia

QoE – Quality of Experience

VQA – Videp Quality Assessment

CODEC – Compression/Descompression

H.264 – Padrão de compressão de vídeo digital também conhecido como MPEG-4 Part 10

MPEG – Moving Picture Experts Group

HVS – Human Visual System

ACR – Absolute Category Rating

DCR – Degration Category Rating

DSIS- Double Stimulus Impairment Scale

FR – Full Reference

RR – Reduced Reference

NR – No Reference

PSNR – Peak Signal to Noise Ratio

MSE- Mean Squared Error

SSIM – Structural Similarity

SLA - Service Level Agreement

AVC – Advanced Video Codec

TV – Television

MB – Macroblocos

IPTV – Internet Protocol Television

UDP – User datagram Protocol

IP – Internet Protocol

TCP – Transmission Control Protocol

UVG – Ultra Video Group

RTMP – Real Time Messaging Protocol

MSU VQMT – MSU Video Quality Measurement Tool

VQM - Video Quality Measurement

NIQE – Natural Image Quality Evaluator

VMAF – Video Multimethod Assessment Fusion

HTML – Hypertext Markup Language

Sumário

Capítulo 1 Introdução	14
1.1 Contextualização do problema.....	14
1.2 Objetivos.....	15
1.2.1 Objetivo geral.....	15
1.2.1.1 Objetivos específicos.....	15
1.3 Organização do trabalho	15
Capítulo 2 Fundamentação Teórica.....	17
2.1 Avaliação da Qualidade de Vídeo (VQA)	17
2.2 Métricas de Avaliação de Qualidade Subjetiva	18
2.3 Métricas de Avaliação de Qualidade Objetiva	19
2.3 Métricas com Referência	21
2.3.1 PSNR	21
2.3.2 SSIM.....	22
2.4 Métricas sem Referência.....	23
2.4.1 Blurring	24
2.4.1 Blocking	24
2.5 Codec	25
2.5.1 Codec H.264.....	26
2.6 Transmissão de vídeo.....	28
2.6.1 Protocolo UDP	29
2.6.1 FFmpeg.....	30
Capítulo 3 Desenvolvimento do Sistema Proposto	31
3.1 Sistema Proposto.....	31

3.2 Vídeos	32
3.3 Transmissão	32
3.4 Avaliação da Qualidade	34
3.4.1 MSU Video Quality Measurement Tool	34
Capítulo 4 Procedimentos Experimentais e Resultados.....	37
4.1 Resultados Vídeo 1: Shake N Dry	37
4.2 Resultados Vídeo 2: Honey Bee	40
4.2 Resultados Vídeo 3: Jockey	43
4.2 Resultados Vídeo 4: Beauty	46
4.5 Análise dos Resultados	49
Capítulo 5 Conclusão	53
5.1 Considerações Finais	53
5.2 Trabalhos Futuros	54
Referências	55

Capítulo 1

Introdução

1.1 Contextualização do problema

Diariamente, cada vez mais conteúdo digital está sendo disponibilizado nas mais diversas plataformas, como redes sociais, streamings, blogs, comércio eletrônico, entre outras. São milhares de vídeos, imagens e textos entregues para os mais diferentes conteúdos e públicos. Dessa forma, se torna essencial para os provedores garantir a qualidade do material entregue e, consequentemente, a satisfação de seus usuários.

Como consequência da grande quantidade de tempo que o olho humano passa observando os conteúdos ofertados digitalmente, os detalhes passam a ser percebidos com maior facilidade e os possíveis defeitos também se tornam mais fáceis de serem encontrados. Assim, é necessário ter conhecimento da chamada Qualidade de Experiência (QoE), para se ter noção da qualidade do material entregue e do contentamento dos usuários. Em relação a vídeos, é aplicada a chamada Avaliação da Qualidade de Vídeo (VQA), que tem como objetivo obter uma perspectiva da qualidade do vídeo entregue.

Existem diversas maneiras de mensurar a VQA de um vídeo, e elas são divididas basicamente em duas: avaliação subjetiva e avaliação objetiva. A avaliação subjetiva utiliza visualizadores humanos que são submetidos a experimentos para analisar a qualidade perceptiva de um conjunto de vídeos. Já a avaliação objetiva utiliza modelos computacionais de estimativa de qualidade e demandam menos esforço e investimentos do que o último.

Para se obter um valor tangível da VQA, são utilizadas métricas de avaliação da qualidade. As métricas de avaliação subjetiva utilizam os visualizadores humanos para qualificar visualmente um conjunto de vídeo, e depois dar uma nota de acordo com determinadas percepções. Já as métricas de avaliação objetiva utilizam os algoritmos e são divididas em três: as de referência completa, que necessitam do vídeo original para ser comparado ao processado; as de referência reduzida, que necessitam somente de algumas informações do vídeo original; e as sem referência, que não necessitam de nenhuma informação do vídeo original.

Para que um vídeo seja entregue, uma série de procedimentos são realizados com o intuito de gravar, processar, transmitir, receber e reproduzir o conteúdo. Uma das etapas mais desafiadoras desse processo é a transmissão, que exige muito do hardware e software utilizados para tal. Para tornar a transmissão de vídeo mais eficiente, é essencial que sejam empregadas técnicas de compressão, que compactam as informações contidas no vídeo diminuindo a quantidade de bits de dados que compõe o arquivo, o que resulta em uma menor demanda de largura de banda para realizar a transmissão. Esse processo é realizado por um par de sistemas, um compressor (codificador) e um descompressor (decodificador), os codecs. O Codec H.264 é um dos mais utilizados com essa finalidade, e é amplamente utilizado em streamings de vídeos, televisão digital, entre outros.

Nesse contexto, esse trabalho visa avaliar a qualidade de experiência em transmissão de vídeo com o codec H.264 utilizando métricas de VQA de referência completa e sem referência. A QoE é essencial para entender como os usuários percebem a qualidade de vídeos transmitidos com o H.264, e a utilização de diferentes tipos de métricas permite uma maior abrangência e confiabilidade nos resultados obtidos.

1.2 Objetivos

1.2.1 Objetivo geral

Realizar a avaliação da qualidade de experiência (QoE) em transmissão de vídeo utilizando o codec H.264 aplicando métricas de avaliação de qualidade de vídeo (VQA) de referência completa e sem referência.

1.2.1 Objetivos específicos

- Realizar a transmissão de amostras de vídeos com o codec H.264;
- Realizar a aplicação de métricas de avaliação de qualidade de vídeo de referência completa e sem referência nas amostras transmitidas;
- Analisar os resultados obtidos através da aplicação das métricas de avaliação de qualidade

1.3 Organização do trabalho

O restante desse trabalho está organizado na seguinte estrutura:

- Capítulo 2: aborda o embasamento teórico e conceitos essenciais para a compreensão do trabalho desenvolvido, como avaliação da qualidade de vídeo, que abrange as métricas de avaliação de qualidade objetiva e subjetiva, o conceito de codec e uma explanação dedicada ao codec H.264, e transmissão de vídeo.
- Capítulo 3: descreve detalhadamente as etapas do projeto desenvolvido para a realização desse trabalho, como a obtenção das amostras de vídeo utilizadas, o processo de transmissão implementado e a aplicação das métricas de qualidade de vídeo empregadas.
- Capítulo 4: apresenta os resultados obtidos da transmissão e da aplicação das métricas de qualidade de vídeo.
- Capítulo 5: apresenta as conclusões finais a respeito dos resultados deste trabalho e trabalhos futuros.

Capítulo 2

Fundamentação Teórica

2.1 Avaliação da Qualidade de Vídeo (VQA)

Com o crescente avanço da tecnologia, também vem crescendo o número de conteúdo ofertado digitalmente. Milhares de textos, imagens e vídeos são disponibilizados todos os dias em streamings, redes sociais, entre outros meios de comunicação. Dessa forma, se torna cada vez mais essencial entregar um conteúdo de qualidade para os consumidores de todas essas informações.

Toda essa exposição do olho humano a materiais cada vez mais elaborados com o intuito de chamar a atenção do telespectador, leva a um desenvolvimento da percepção do usuário, principalmente no quesito qualidade do conteúdo. Logo, essas considerações levantam a importância de os provedores de serviços conhecerem a Qualidade da Experiência (QoE) dos usuários e os usuários conhecerem a qualidade recebida. (ROSSHOLM et al., 2014)

No contexto de vídeos, existe a chamada Avaliação da Qualidade de Vídeo (VQA), que se refere a um processo com o intuito de estimar a qualidade perceptiva de um vídeo. (PANDREMMENOU et al., 2014) Durante o processo de gravação, processamento, transmissão, recepção e reprodução do vídeo, podem ocorrer várias distorções que eventualmente acarretam na entrega de um material de má qualidade ao usuário final. Para evitar que isso aconteça, há um grande investimento na VQA por parte dos provedores.

Existem vários métodos que podem ser utilizados para mensurar a qualidade de um vídeo. Eles são basicamente divididos em dois: a avaliação subjetiva (que usa visualizadores humanos) e a avaliação objetiva (realizada através de algoritmos).

A avaliação subjetiva se dá por meio de experimentos que envolvem um painel de participantes que são geralmente não especialistas, também chamados de sujeitos de teste, para avaliar a qualidade perceptiva de determinado material de teste, nesse caso um conjunto de vídeos. (ROSSHOLM et al., 2014) Para se realizar um experimento subjetivo é necessário uma série de cuidados, tanto com os integrantes quanto com o material e ambiente utilizados. É fundamental um ambiente controlado, com determinadas condições de visualização que depende do experimento, uma seleção detalhada do material apresentado, elaboração do

método de avaliação, que geralmente incluem escalas numéricas, e controle do tempo de apresentação. Esse conjunto de requisitos tornam esse tipo de avaliação cara, demorada e cansativa.

Por outro lado, a avaliação objetiva geralmente consiste em modelos computacionais de estimativa de qualidade e podem ser automatizados, são considerados eficientes em termos de tempo e mais adequados para aplicações em tempo real. (ROSSHOLM et al., 2014) Esses métodos tem como objetivo avaliar a qualidade do vídeo o mais próximo possível de determinada característica da percepção humana. Dentro da avaliação objetiva, existem diversos tipos de métodos que podem ser classificados de diferentes maneiras dependendo da abordagem utilizada. Pode-se utilizar, por exemplo, uma comparação entre o vídeo original e o vídeo transmitido para se obter uma avaliação da qualidade do material recebido.

2.2 Métricas de Avaliação de Qualidade Subjetiva

Durante os processos que envolvem os sistemas de vídeo, podem ocorrer várias percas e distorções no conteúdo ofertado ao usuário. Dessa forma, os provedores geralmente fazem compensações em componentes e elementos de seu sistema de modo a garantir a máxima qualidade possível em seus serviços.

A forma mais evidente e antiga utilizada para mensurar a qualidade de um vídeo, é coletando a opinião do telespectador diante de determinado material. Logo, as métricas de avaliação de qualidade subjetiva utilizam humanos para qualificar visualmente um vídeo ou um conjunto de vídeos, com base em percepções e sensibilidades óticas, através de experimentos manipulados.

As métricas subjetivas são influenciadas por vários motivos, entre eles: participantes, o número de visualizadores dos vídeos varia muito e depende do tipo de abordagem utilizada; tipo de dispositivo: os dispositivos utilizados no experimento influenciam diretamente na experiência do usuário, pois diferentes equipamentos possuem diferentes especificações e resoluções que impactam na exibição do material; Sistema Visual Humano (HVS): o sistema de visão dos participantes também pode impactar os resultados, visto que enfermidades oculares são uma possibilidade e podem interferir na avaliação; condições de visualização: o ambiente em que é realizado o experimento também tem importante contribuição nos resultados, uma vez que fatores como iluminação e distancia da tela são essenciais para a experiência do visualizador; método utilizado para avaliação: as métricas subjetivas utilizam diferentes critérios para determinar a qualidade de um vídeo, que influenciam diretamente no

resultado; assuntos: como os humanos são o objeto principal das métricas subjetivas, é necessário levar em consideração os sentimentos e emoções dos mesmos, visto que os visualizadores tendem a dar uma maior classificação a aqueles vídeos com conteúdo que mais os agrada. Esses múltiplos fatores podem influenciar significativamente a avaliação subjetiva. (NIDHI; AGGARWAL, 2014)

Uma das métricas subjetivas mais utilizadas é a chamada Absolute Category Rating (ACR). Esse método leva em consideração várias categorias na avaliação do visualizador, com base em sua percepção pessoal. Os vídeos são apresentados de forma individual, um de cada vez, e ao final da exibição o participante é indicado a dar uma nota. A métrica ACR utiliza a seguinte escala de classificação em cinco níveis, conforme exibido na Tabela 1. (AKRAMULLAH, 2014)

Tabela 1 – Classificação ACR

Valor	Classificação
5	Excelente
4	Bom
3	Razoável
2	Baixo
1	Ruim

Fonte: Adaptado de (AKRAMULLAH, 2014).

Outra métrica subjetiva bastante aplicada é a Degradation Category Rating (DCR), também chamada de Double Stimulus Impairment Scale (DSIS). Nesse tipo de métrica, é exibido ao visualizador o vídeo original e logo em seguida o vídeo distorcido, e então é solicitado uma classificação referente ao segundo conteúdo, baseada na comparação com o material de referência. A escala classificatória para o DCR está exibida na Tabela 2. (NEHMÉ et al., 2020)

Tabela 2 – Classificação DCR

Valor	Classificação
5	Imperceptível
4	Perceptível, mas não incômodo
3	Pouco incômodo
2	Incômodo
1	Muito incômodo

Fonte: Adaptado de (AKRAMULLAH, 2014).

2.3 Métricas de Avaliação de Qualidade Objetiva

Nos últimos tempos houve um grande investimento em VQA que visam o desenvolvimento de modelos e algoritmos que sejam capazes de qualificar um vídeo da forma mais compatível possível com a percepção humana. Dessa forma, é poupado tempo e energia com métodos que apresentam resultados quase que instantâneos e que podem ser automatizados para repetirem quantas vezes necessário. Dessa forma, modelos computacionais de VQA foram projetados e geralmente são denominados métodos objetivos (alternativamente, métricas ou modelos). (ROSSHOLM et al., 2014)

Vale ressaltar que por mais que os algoritmos com o objetivo de avaliar a qualidade de vídeos em sua maioria são desenvolvidos com o intuito de corresponder ao julgamento humano, os resultados das métricas nem sempre obtém resultados pertinentes quando comparados a assimilação do homem. Assim, a avaliação objetiva da qualidade não pode substituir a avaliação subjetiva da qualidade. Elas apenas auxiliam como uma ferramenta na avaliação da qualidade. (AKRAMULLAH, 2014)

As métricas objetivas de qualidade de vídeo podem ser desenvolvidas com base em duas abordagens diferentes: psicofísicas e de engenharia. As métricas com abordagem psicofísicas têm como padrão o sistema visual humano (HVS) para desenvolver um padrão de análise de qualidade. São utilizados sentidos e percepções humanas como sensibilidade de contraste e captação de cores como fundamentos para o desenvolvimento dos algoritmos. Já a abordagem de engenharia usa métricas simplificadas com base na extração e análise de certas características ou artefatos em um vídeo, mas não necessariamente desconsidera os atributos de o HVS. (PANDREMMENOU et al., 2014) Entretanto, seu foco é averiguar o conteúdo do vídeo e suas distorções, e não considerar todos os aspectos da visão humana. Um conjunto de aspectos e propriedades associados a qualidade de vídeo são reunidos para que um método de avaliação objetiva seja criado na abordagem de engenharia.

As métricas objetivas de qualidade de vídeo podem ser aplicadas com diversos propósitos e em diferentes contextos. Elas são muito úteis quando se é necessário fazer a avaliação da qualidade de um vídeo em tempo real, e realizar certas adequações para que essa qualidade se torne apropriada para o tipo de transmissão que está sendo realizada. Outra aplicação é quando se deseja melhorar os padrões de transmissão e recepção, pois com base nos resultados obtidos é possível realizar ajustes para que a qualidade do vídeo seja ainda

melhor. Elas também podem ser aplicadas a modo de comparação, entre diferentes processamentos e sistemas de transmissão, por exemplo.

Quanto a classificação das métricas objetivas, elas podem ser divididas de acordo com o nível de informação do vídeo original que exigem. Essa especificação é segmentada em basicamente três tipos: referência completa (FR), referência reduzida (RR) e sem referência (NR).

As métricas FR são aquelas que tem a necessidade do vídeo original completo para ser usada como referência. Dessa forma, elas fazem uma comparação entre o arquivo original e o processado em busca de elementos divergentes. As métricas RR não precisam do vídeo completo, mas somente de algumas informações sobre ele. Geralmente são utilizadas características determinantes, como intensidade, profundidade e texturas. Já as métricas NR não necessitam de nenhuma informação a respeito do vídeo original. Assim, elas analisam somente o vídeo processado com base em diversos outros parâmetros. Logo, as métricas NR podem apresentar resultados não tão confiáveis como as FR, mas são muito úteis quando o vídeo original não está disponível.

2.4 Métricas com Referência

As métricas de referência completa necessitam do vídeo original disponível, que é usado como parâmetro de comparação com o vídeo distorcido. Dessa forma, é feita uma verificação da semelhança das características dos dois conteúdos, com o intuito de obter a qualidade do vídeo processado.

Em comparação com as métricas RR e NR, as métricas FR são as que possuem melhor desempenho. Esse fato se deve em grande parte com a disponibilidade do vídeo original que pode ser usado como base de referência, gerando um resultado muito mais confiável. Entretanto, a necessidade de referência também pode ser um empecilho, visto que em alguns casos isso não é possível, como por exemplo em transmissões de tempo real. Além disso, é necessário um grande esforço para alinhar os dois vídeos, visto que os frames devem estar perfeitamente igualados para garantir a precisão da métrica. (FARIAS, 2010)

2.4.1 PSNR

A Relação Sinal-Ruído de Pico, ou do inglês, Peak Signal to Noise Ratio é uma métrica de referência completa utilizada com o objetivo de obter a qualidade de um vídeo. Ela pode ser

definida como a relação entre a máxima energia possível de um sinal e a energia do ruído que danifica a representação entre os frames do vídeo original e o processado.

O PSNR é determinado por meio do cálculo do Mean Squared Error (MSE), ou Erro quadrático médio. O MSE é um medidor da diferença média quadrada entre o valor estimado e o parâmetro. Ele especifica a diferença média dos pixels em todos os frames do vídeo original com o processado. Quanto maior for o valor do MSE, maior a distorção presente entre os vídeos. Ele é dado pela equação apresentada a seguir. (POOBATHY; CHEZIAN, 2014)

$$MSE = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} [f(x, y) - g(x, y)]^2$$

Onde, $f(x,y)$ é a variável de entrada (valor da cor do pixel original), $g(x,y)$ é a variável de saída (valor do pixel processado), m é o número de pixels na direção horizontal e n é o número de pixels na direção vertical. Com o MSE definido, é possível dar a equação que descreve o PSNR em si.

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right)$$

Onde R é a variação máxima nos dados de imagem de entrada. Para uma representação de n bits do sinal, $R = 2^n - 1$. Considerando um caso onde um sinal possui 8 bits, então $R = 255$. Outro exemplo, seria um sinal possuir 10 bits, logo $R = 1023$. O valor resultante para o PSNR é expresso em termos de escala logarítmica de decibéis (dB).

Considerando que o PSNR é uma das métricas mais simples e de fácil mensurabilidade, ela é amplamente utilizada para avaliar a qualidade vídeos e imagens. Uma de suas aplicações mais comuns, é em comparações de codificadores de vídeo, como objetivo de determinar qual deles é o que menos impacta em termos de qualidade. Outras utilizações do PSNR são: a avaliar a compressão de vídeos e algoritmos de processamento, e identificar os frames corrompidos ou perdidos em uma sequência de vídeos. (AKRAMULLAH, 2014)

2.4.2 SSIM

Algumas métricas de qualidade objetiva utilizam matemática pura de cálculo de erro para realizar suas medições, como o PSNR. Entretanto, outros fatores podem fornecer uma melhor estimativa da qualidade. O Sistema Visual Humano (HVS) é capaz de extrair complexas informações da estrutura visualizada, e essa característica pode ser bastante útil em métricas

que utilizam como base elementos estruturais para realizar suas avaliações, entregando um resultado muitas vezes melhor do que as que usam a sensibilidade ao erro. (DAVCEV; GÓMEZ, 2009) Esse fato se dá pelo HVS não ser especialista em detectar e extrair erros, mas sim em processar as informações captadas dentro de seu campo de visão.

A métrica chamada Similaridade Estrutural, ou do inglês Structural Similarity, SSIM utiliza exatamente essa abordagem. O índice de Similaridade Estrutural é definido pela combinação de três componentes da imagem: comparação de luminância, comparação de contraste e comparação de estrutura. A equação matemática do SSIM é apresentada a seguir.

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

Onde x e y são janelas sendo comparadas de dimensões $N \times N$, μ_x é a média de x , μ_y é a média de y , σ_x^2 é a variância de x , σ_y^2 é a variância de y , σ_{xy} é a covariância de x e y , $C_1 = (k_1L)^2$ e $C_2 = (k_2L)^2$ são duas variáveis para estabilizar a divisão, L é a faixa dinâmica dos valores dos pixels e geralmente é $2^{bitsporpixel} - 1$, $k_1 = 0.01$ e $k_2 = 0.03$.

2.5 Métricas sem Referência

As métricas sem referência realizam a análise somente do vídeo distorcido ou processado, sem a necessidade de ter o arquivo original como parâmetro. Dessa forma, são adequadas para uso em alguns casos específicos onde não há acesso ao vídeo original, como por exemplo quando é necessário avaliar o desempenho da qualidade em tempo real durante transmissões.

Considerando os investimentos feitos às métricas sem referência recentemente, as aplicações para esses tipos de métodos são muitas, e incluem principalmente os serviços multimídia. Algumas das aplicações das métricas NR são: prover a quantificação da qualidade do serviço entregue aos usuários de provedores de conteúdo, de modo que possa ser feito ajustes e melhorias na arquitetura do sistema de transmissão para fornecer a melhor experiência; Com o envolvimento de vários provedores de conteúdo e também de usuários, foi necessário estabelecer o acordo de nível de serviços (SLA), que tem como objetivo definir um certo nível de qualidade que deve ser entregue ao cliente. Uma forma de garantir que esse nível seja alcançado é através das medições da qualidade utilizando as métricas NR; Em sistemas de transmissão de tempo real é necessário verificar a qualidade do vídeo transmitido e a partir daí fazer adaptações para garantir uma boa entrega. Nesse caso, as métricas de referência têm

aplicação limitada pois os vídeos originais dificilmente estão disponíveis, fazendo com que as métricas sem referência sejam as mais indicadas para ajustes em simultâneos. (PANDREMMENOU et al., 2014)

2.5.1 Blurring

O desfoque, ou blur do inglês, pode ser definido como um artefato que resulta na perda de detalhes e danos na resolução nas bordas ou em áreas texturizadas de imagens e quadros de vídeos. (WINKLER, 2005) Os motivos para o desfoque ocorrer são variados, e podem acontecer em diversos estágios de sistema de vídeos, como na gravação, processamento e transmissão. Algumas das causas mais comuns do desfoque são a perda do foco da captura, o movimento da câmera e do objeto que está sendo registrado, limitações no sistema ótico de captura. (SHAHID et al., 2014)

Dessa forma, a métrica de avaliação de qualidade Blurring estima a variação de cor entre pixels vizinhos e calcula uma variação média. Geralmente são aplicados métodos para reconhecer a borda dos objetos presentes no quadro de vídeo, e estes frames são os avaliados primeiramente. Após isso, é finalmente obtido a medida de desfoque para toda a imagem, por meio do cálculo da média dos valores de desfoque local em todos os quadros de borda.

Figura 1 – Imagem original e imagem com o artefato blurring.



Fonte: Adaptado de (RICHARDSON, 2010).

2.5.1 Blocking

O artefato de bloqueio, ou do inglês blocking, pode ser definido como uma descontinuidade entre blocos adjacentes de imagens e blocos de vídeos. (WU; RAO, 2005) Esse

efeito é considerado um dos mais desagradáveis visualmente quando se trata de qualidade. O bloqueio é frequentemente causado no processamento baseado em blocos, pela quantização grosseira de cada bloco individualmente, o que ocasiona em limites de blocos defeituosos nas imagens ou quadros no momento da reconstrução. Portanto, o bloqueio é um artefato que se deve evitar ao máximo em sistemas de vídeos e imagens.

Dessa forma, a métrica de avaliação de qualidade blocking é usada para estimar o bloqueio em uma região de uma imagem ou frame de vídeo, geralmente por meio de cálculos da diferença entre blocos vizinhos e a quantidade de brilho ao redor desses blocos. Frequentemente, a detecção dos bloqueios leva os seguintes aspectos em consideração: a intensidade do bloqueio (a diferença perceptiva entre os blocos adjacentes), o brilho ao redor dos blocos e a textura presente. (SHAHID et al., 2014)

O blocking utiliza diversos métodos para detectar as bordas de objetos, que são consideradas como as bordas dos blocos. Desse modo, a métrica é aplicada somente nesses pontos e seu valor é reduzido, o que permite mensurar o bloqueio com maior precisão. Podem ser considerados blocos de imagens de várias dimensões, porém a mais comumente aplicada é a 8x8. (VQMT Metrics, 2023)

Figura 2 – Imagem original e imagem com o artefato blocking.



Fonte: Adaptado de (RICHARDSON, 2010).

2.6 Codec

No mundo digital, a compressão pode ser definida como o ato de representar informações de forma compactada. Esse processo é feito com o intuito de diminuir o número de bits de dados que compõe o arquivo. A compressão de vídeos, mais especificamente, é

realizada em vídeos digitais para diminuir o número de bits, e assim facilitar o processo de transmissão ou armazenamento.

O processo de compressão é composto por um par de sistemas, um compressor (codificador) e um descompressor (decodificador). A função do codificador é reduzir o número de bits dos dados de origem, antes da transmissão e armazenamento, enquanto o decodificador transforma esses dados compactados de volta na informação original. Esse par de sistemas é descrito como CODEC, do inglês encoder (codificador) e decoder (decodificador). (RICHARDSON, 2010)

A compressão de dados pode ser dividida em dois tipos, considerando a forma como a reconstrução da informação é realizada. O primeiro deles é o esquema de compressão sem perdas, onde a informação de entrada é igual a informação de saída. O segundo é o esquema de compressão com perdas, onde geralmente o poder de compactação é muito maior, o que acarreta em informações de entrada e saída diferentes.

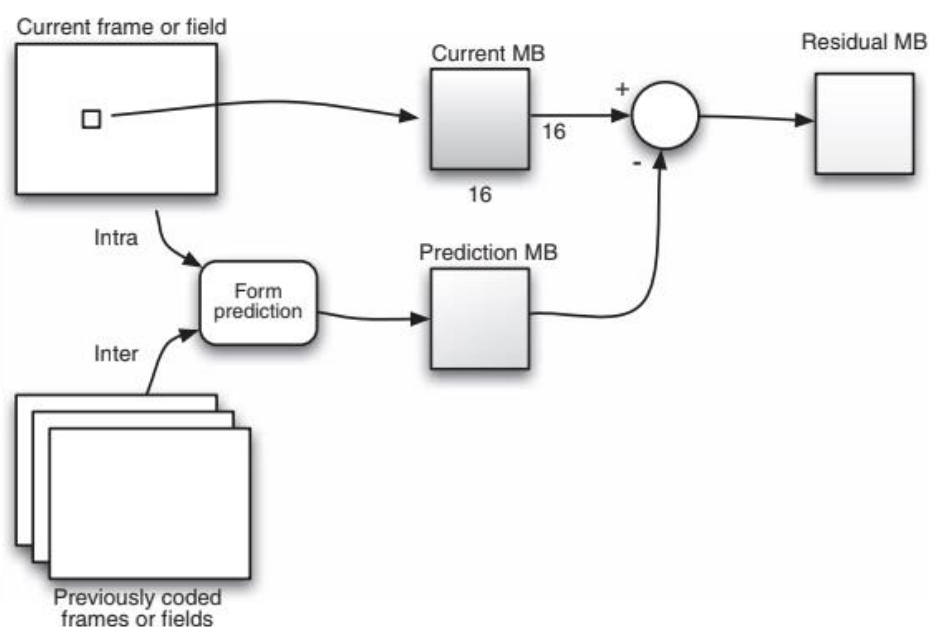
Quando se trata de compressão de vídeos, os sistemas são projetados com o intuito de remover informações redundantes, que não comprometeriam a percepção do espectador. A maior parte dos métodos de codificação de vídeos leva em consideração redundâncias temporais e espaciais, no momento de eliminar alguns dados. (RICHARDSON, 2010) As redundâncias espaciais ocorrem quando um único quadro dentro de uma sequência de vídeo contém uma quantidade notável de espaço redundante, ou seja, os pixels estão muito próximos uns dos outros, e são muito semelhantes. Nesse caso, o processo de compressão ocorre de maneira a eliminar a redundância existente em cada quadro. Já as redundâncias temporais ocorrem quando um grupo sucessivo de quadros de um vídeo são muito semelhantes, o que ocorre geralmente por uma elevada taxa de quadros. Dessa forma, para realizar a compressão, somente a diferença entre os quadros é enviada. Isso significa que somente parte do quadro que sofreu mudanças em relação ao anterior é enviada. (CHEN et al.)

2.6.1 Codec H.264

O H.264, também chamado de Advanced Video Codec (AVC), é um método e formato para compressão de vídeos, um processo que consiste em realizar a conversão de um vídeo digital em um formato menor, e que exija menos da transmissão ou armazenamento. A compressão de informações, em especial a de vídeos, é essencial para o funcionamento de streamings, TV digital, videoconferências, entre outros meios de comunicação. (RICHARDSON, 2010)

Durante o processo de codificação, o H.264 realiza processos de predição, transformação e codificação para compactar um vídeo, gerando um fluxo de bits bem menor. A partir daí, durante o procedimento de decodificação, o H.264 realiza processos de decodificação, transformação inversa e reconstrução para que seja possível reproduzir a informação decodificada.

Figura 3 – Processo de codificação



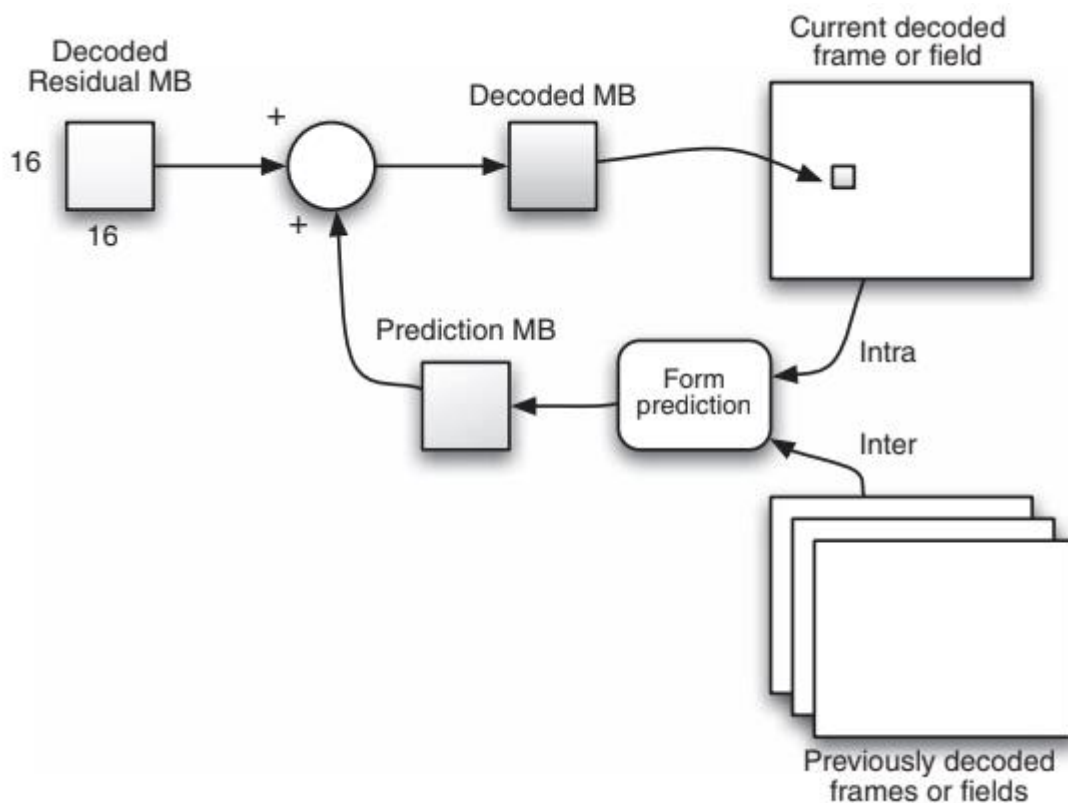
Fonte: Richardson, I. (2010)

O H.264 utiliza a codificação por predição para produzir um fluo de bits compactado. Esse processo tenta encontrar blocos de pixels já codificados que sejam semelhantes a blocos que estejam presentes no quadro atual, para que ao invés de todo o quadro atual seja processado, somente a sua diferença precise ser codificada. (CHEN et al.) Dessa forma, a informação decodificada geralmente não é idêntica a original, pois o H.264 realiza uma compactação com perdas e, portanto, alguma parte da qualidade é perdida durante esse processo.

Esses blocos de pixels são os chamados macroblocos (MB) e correspondem a 16x16 pixels exibidos. Durante o processo de codificação, um macrobloco de previsão é gerado e subtraído do macrobloco original para formar um macrobloco residual, que será quantizado e codificado. Em paralelo, esse macrobloco residual é transformado inversamente e adicionado ao macrobloco de previsão para auxiliar na reconstrução de quadros posteriores. No processo de decodificação, o macrobloco é decodificado para gerar o macrobloco residual decodificado.

O decodificador também era uma previsão e adiciona ao macrobloco residual decodificado, para produzir um macrobloco decodificado. (RICHARDSON, 2010)

Figura 4 – Processo de decodificação



Fonte: Richardson, I. (2010)

2.7 Transmissão de vídeo

Quando se trata de transmissão de vídeos, existe uma ampla variedade de diferentes aplicações e implementações operacionais com diversos propósitos. A transmissão do vídeo pode ser realizada com o objetivo de comunicação ponto a ponto, *multicast* ou *broadcast*, o vídeo pode ser pré codificado, armazenado, ou codificado em tempo real, uma videoconferencia por exemplo, entre muitas outras variáveis. O canal por onde essa transmissão será realizada também possuem diversas possíveis configurações, como estáticos ou dinâmicos, comutados por pacotes ou circuito, entre outros. (APOSTOLOPOULOS et al., 2002)

O modelo de transmissão de vídeo mais popular é a chamada comunicação one-to-all, do inglês um para todos, ou comunicação *broadcast*, onde basta apenas uma fonte de transmissão para todos que possam acessar o conteúdo disponibilizado. O exemplo mais conhecido desse tipo de transmissão é a transmissão de televisão. Uma característica importante

da comunicação one-to-one é a impossibilidade de retorno para o remetente, devido ao grande número de receptores, o que torna inviável a adaptação e melhoria do sistema.

Outro tipo de transmissão é a comunicação unicast ou one-to-one, ou um a um, onde a transmissão ocorre unicamente entre dois pontos. Um exemplo desse tipo de transmissão é a videochamada individual. Já a comunicação multicast é uma transmissão que fica entre a broadcast e a unicast, pois sua característica é a comunicação one-to-many, ou seja, um para muitos. Um exemplo de comunicação multicast é o IPTV (Internet Protocol Television), onde vários usuários acessam a mesma informação ao mesmo tempo. (APOSTOLOPOULOS et al., 2002)

A transmissão de vídeo também pode ter de duas formas diferentes quando analisada do lado receptor. A primeira forma é baseada em pré codificação e download, no qual o arquivo de vídeo é baixado completamente para o armazenamento local antes da reprodução. Nesse cenário, o tempo necessário para baixar um vídeo cresce à medida que a quantidade de dados aumenta, o que está diretamente relacionado à qualidade e a duração do vídeo. Além disso, a largura de banda da rede desempenha um papel crucial do tempo de download. A segunda forma é a codificação em tempo real, onde o arquivo é visualizado ao mesmo tempo em que a transmissão é realizada. (SUN et al., 2007)

2.7.1 Protocolo UDP

User Datagram Protocol (UDP) é um protocolo de rede utilizado para transmissão de vídeos, que assume o IP (Internet Protocol) como protocolo subjacente. (POSTEL, 1980) O IP proporciona a entrega de rede com o menor esforço, fornecendo endereçamento, roteamento e um formato global que possibilita a interpretação por todos. No topo do IP, estão os protocolos de transporte ponto a ponto, entre eles o UDP. (APOSTOLOPOULOS et al., 2002) Ele surgiu como uma alternativa ao TCP (Transmission Control Protocol), um outro protocolo de transporte considerado mais seguro e confiável por realizar a transmissão através do uso de firewall, um componente de segurança de rede. Ele também garante que a entrega foi realmente realizada por confirmações e retransmissões.

O UDP não é considerado tão seguro quanto o TCP por simplesmente proporcionar uma interface de IP para o usuário, sem nenhuma verificação de segurança. O UDP não é capaz de garantir que a entrega foi realizada com sucesso, pois ele permite que os pacotes sejam descartados caso estiverem danificados, portanto o usuário poderá visualizar o vídeo danificado, mas o fluxo continuará.

2.7.2 FFmpeg

FFmpeg é uma solução completa e multiplataforma que permite a gravação, conversão e transmissão de áudio e vídeo. Ele consiste em um conjunto de ferramentas e linhas de comando que podem ser usadas para gravar, converter e transmitir áudio e vídeo digital em vários formatos. (NEWMARCH, 2017) A rápida velocidade de processamento fez com que a ferramenta se tornasse amplamente conhecida e utilizada no ramo do processamento de sinais.

O FFmpeg é um software livre licenciado pela GNU General Public License. “FF” de FFmpeg significa Fast Forward, que é um botão de controle em reprodutores de mídia, “mpeg” é uma abreviação de Moving Pictures Experts Group. A linguagem de programação utilizada no FFmpeg é C, e o código fonte pode ser compilado em Linux, Windows, Mac, OS entre outros. (KORBEL, 2012)

Algumas ferramentas das linhas de comandos do FFmpeg são: (KORBEL, 2012)

- ffmpeg: codificação e decodificação rápida de áudio e vídeo.
- ffplay: reprodutor de mídia.
- ffprobe: mostra as características dos arquivos de mídia.
- ffmpegserver: servidor de transmissão para streaming multimídia usando protocolos HTTP e RTSP;

Algumas bibliotecas do FFmpeg são: (KORBEL, 2012)

- libavcodec: biblioteca de software para vários codecs multimedias
- libavdevice: biblioteca de software para dispositivos
- libavfilter: biblioteca de software contendo filtros
- libavformat: biblioteca de software para formatos de mídia
- libavutil: biblioteca de software contendo vários utilitários
- libpostproc: biblioteca de software para pós processamento
- libswresample: biblioteca de software para amostragem de áudio

Capítulo 3

Desenvolvimento do Sistema Proposto

3.1 Sistema Proposto

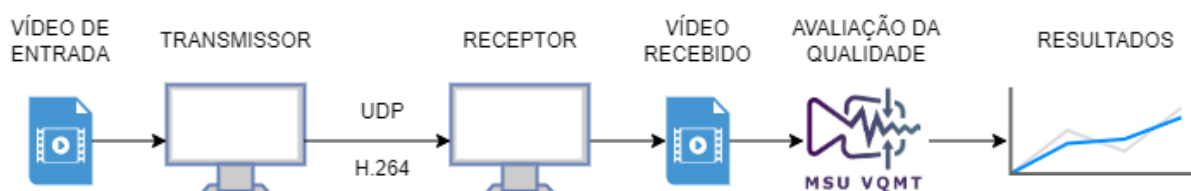
O desenvolvimento do sistema desse projeto consiste em três etapas: obtenção dos vídeos, transmissão dos vídeos e avaliação da qualidade dos vídeos. Como entrada do sistema, foram utilizadas sequências de vídeos obtidas através de um conjunto de dados abertos chamado Ultra Video Group. Foram escolhidas quatro sequências distintas com diferentes fundos, para uma melhor amostra de resultados obtidos. Os vídeos obtidos do conjunto foram disponibilizados em formato H.264/AVC.

Com as sequências obtidas, foi necessário realizar a transmissão dos vídeos. Para o sistema de transmissão, foram utilizados dois computadores, um como emissor e outro como receptor, e o a suíte de software FFmpeg. Foram utilizados linhas de comandos para transmitir o vídeo do computador emissor, e para receber e salvar sem nenhum tipo de processamento no computador receptor.

Uma vez que os vídeos foram transmitidos e salvos, foi preciso realizar a avaliação da qualidade dos mesmos. Para essa etapa foi utilizado o software MSU Video Quality, que possui diversas métricas sem referência e com referência disponível que podem ser aplicadas em sequências de vídeos com o objetivo de obter um valor mensurável da qualidade do vídeo em questão. Foram escolhidas quatro métricas, duas com referência completa e duas sem referência, para uma melhor amostragem dos resultados.

Finalmente, com o valor das métricas aplicadas aos vídeos que foram transmitidos, foi possível comparar os resultados obtidos e fazer uma avaliação da qualidade dos vídeos em formato H.264 após passar pelo processo de transmissão.

Figura 5 – Diagrama do sistema proposto.



3.2 Vídeos

Para a realização desse projeto, foi necessário a obtenção de vídeos que foram utilizados como amostras para o sistema proposto. Os vídeos foram obtidos a partir do conjunto de dados abertos do Ultra Vídeo Group (UVG). (MERCAT; VIITANEN; VANNE, 2020) Ele é mantido por um grupo de pesquisa da unidade de Ciências da Computação da Universidade de Tempere, na Finlândia. Nesse conjunto de dados abertos, é possível acessar dezesseis versáteis sequências de vídeos de teste 4K (3840 x 2160) capturadas a 50/120 fps.

Das dezesseis sequências disponíveis foram escolhidas quatro, chamadas: ShakeNDry, HoneyBee, Jockey e Beauty. A sequência ShakeNDry exibe um cão correndo em uma floresta. A sequência HoneyBee mostra uma abelha voando entre várias flores. A sequência Jockey exibe um cavalo em uma pista de corrida. A sequência Beauty mostra o rosto de uma mulher em um fundo preto.

É possível ainda escolher alguns parâmetros da sequência de vídeos, como resolução, formato e profundidade de vídeo. Para este projeto, foram escolhidas as seguintes características: resolução 4K, profundidade de bits igual a 8, codec AVC/H.264 e contêiner MP4. A resolução se refere a quantidade de detalhes no vídeo ou nitidez, e é a medida em pixels na largura e altura de uma imagem de vídeo. A profundidade de bits faz referencia a quantidade de bits que são utilizados para representar a cor ou tons de cinza em um pixel. Codec faz referencia a forma de como o vídeo é codificado e decodificado. O contêiner é um tipo de formato que agrupa os arquivos de som e imagem.

3.3 Transmissão

Para realizar o processo de transmissão do sistema proposto, foram utilizados dois notebooks. O notebook de emissão foi um modelo Samsung Book, com processador Intel Core i5, placa de vídeo Intel Iris Xe Graphics e Windows 11. O notebook de recepção foi um modelo Lenovo Ideapad 3i, com processador Intel Core i7, placa de vídeo NVIDIA GeForce MX330 e Windows 11. A transmissão foi realizada em ambiente indoor, com os dois notebooks separados por uma distância de dois metros um do outro.

Para a transmissão dos vídeos em si, foi utilizada a suíte FFmpeg. A mesma é uma plataforma que permite a conversão de formatos, gravação e edição de áudio e vídeo, e streaming de mídia por meio de vários formatos diferentes, entre eles RTMP (Real-Time

Messaging Protocol) ou por outros meios e servidores. A forma mais viável e escolhida foi a transmissão UDP (User Datagram Protocol).

Para a transmissão UDP ocorrer, foi necessário primeiramente instalar o FFmpeg em ambas os computadores usados no projeto. Para isso, foi acessado o site oficial do FFmpeg, e realizado o download da versão “Windows” 64 bits. Como o arquivo é baixado de forma compactada, foi necessário realizar a descompactação do mesmo. Após o arquivo descompactado, o FFmpeg foi adicionado ao PATH de ambos os sistemas, para que dessa forma, fosse possível o uso da suíte a partir do Prompt de Comando do Windows. Com todas essas etapas concluídas,

Para o processo de transmissão dos vídeos, foi utilizado o comando FFmpeg exibido no Código 1. Esse comando serve para transmitir um arquivo de vídeo local, através da transmissão UDP, com o codec H.264. A parte “-re” do comando instrui o FFmpeg a ler o arquivo de entrada. A transmissão é realizada com repetição infinita, para que fosse viável o salvamento no momento da recepção, através da parte “-stream_loop -1” do comando. A parte “-i <local do arquivo>” especifica o arquivo de entrada a ser transmitido e sua localização. A parte “-c:v libx264” define o codec H.264 para o arquivo de saída. A taxa de bits foi definida para 3000 kbps (kilobits por segundo), o que auxilia na qualidade da compressão do vídeo. Outras partes importantes do comando são “-preset ultrafast” e “-tune zerolatency”, que definem, respectivamente, a velocidade de codificação para “ultrafast” e a latência pra o mínimo possível, auxiliando a transmissão ao vivo. A parte “an” do comando significa que o vídeo será transmitido sem áudio. “-f mpegts” especifica o formato de saída como MPEG-TS, que é comumente usado em transmissões UDP. “udp://<ip de destino>:<porta>” significa que a transmissão será feita com o protocolo UDP para o IP e porta de destino especificadas. Foi preciso modificar essas duas informações para o IP do computador receptor e uma porta que estivesse livre no momento tanto quanto para o computador receptor, quanto para o computador transmissor.

Código 1 – Comando FFmpeg utilizado para transmissão.

```
ffmpeg -re -stream_loop -1 -i <local do arquivo>-c:v libx264 -b:v 3000k -preset ultrafast -  
tune zerolatency -na -f mpegts udp://<ip de destino>:<porta>
```

Para a receber e salvar os vídeos, foi utilizado o comando FFmpeg exibido no Código 2. Esse comando permite receber mídias, no caso um vídeo, através da transmissão UDP e criar uma cópia local. “-i udp://<ip de origem>:<porta>” significa a fonte de entrada de onde o

FFmpeg deve capturar o fluxo de vídeo e áudio, no caso uma comunicação via protocolo UDP. As partes “-c:v copy” e “-c:a copy” do código permitem que vídeo nem o áudio sejam reprocessados ou convertidos no processo de recepção e salvamento, ou seja, o vídeo não sofre nenhum processamento ou conversão, ele é copiado diretamente. “output.mp4” especifica o nome do arquivo MP4 que será salvo.

Código 2 – Comando FFmpeg utilizado para recepção.

```
ffmpeg -i udp://<ip de origem>:<porta> -c:v copy -c:a copy output.mp4
```

3.4 Avaliação da Qualidade

Com os vídeos transmitidos e armazenados, foi possível partir para a terceira e última etapa do projeto.

Para a realização da avaliação de qualidade de vídeo (VQA), foi utilizado o software gratuito MSU Video Quality Measurement Tool (MSU VQMT), que é uma ferramenta para análise de qualidade de vídeo e imagem utilizando métricas de referência completa (FR) e sem referência (NR).

3.4.1 MSU Video Quality Measurement Tool

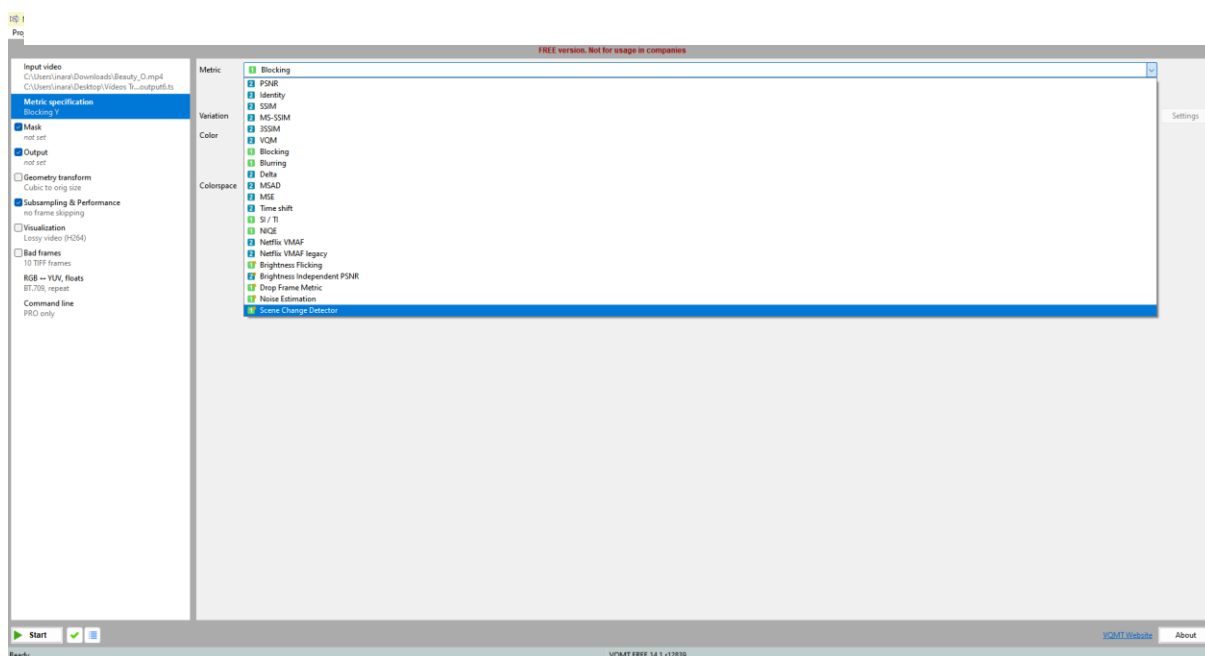
O MSU VQMT, sigla para Ferramenta de Medição de Qualidade de Vídeo MSU em inglês, é um software profissional que tem como proposta realizar análises profundas e objetivas da qualidade de vídeo. Sua principal funcionalidade consiste em calcular métricas objetivas de qualidade para conteúdos de vídeo e imagem. Esses cálculos podem ser realizados tanto por meio de análises com referência, onde uma sequência de vídeo processada ou compactada é comparada com a original, quanto por meio de análises sem referência, em que o conteúdo e sua qualidade são avaliados sem comparações. O software é desenvolvido pela equipe do COMPRESSION.RU com participação do MSU Graphics & Media Lab. Video Group (G&M Lab) e está disponível em diversas licenças, entre elas a versão Free. MSU VQMT Brief Description. Disponível em: <<https://videoprocessing.ai/vqmt/basic/>>. Acesso em: 25 de outubro de 2023.

O MSU VQMT tem abrangentes aplicações, dentre elas estão o uso por desenvolvedores de codecs, indústria televisiva, grupos de pesquisa e universidades, provedores de vídeo, fabricantes de câmeras, compressores de vídeo e transmissores, entre muitos outros.

Essa ferramenta suporta diversos tipos de métricas de qualidade de vídeo, tanto com referência, como sem referência. Algumas delas são: PSNR, MSE, SSIM, VQM, NIQE e VMAF.

A Figura 6 mostra a interface da ferramenta. No lado esquerdo, é possível observar o menu de configuração, onde temos as duas principais especificações a definir: Input vídeo, onde é inserido o caminho do vídeo a ser analisado, e Metric specification, onde é definida a métrica que será aplicada para a avaliação da qualidade de vídeo. As métricas que possuem o número 1, são as sem referência. As métricas com o número 2, são as com referência. Caso seja escolhida uma métrica com referência, é possível inserir o caminho do vídeo original, e do vídeo processado no menu Input vídeo.

Figura 6 –Interface do MSU VQMT



Após dois parâmetros configurados, é possível aplicar a medição da qualidade através do cálculo da métrica escolhida. Para isso, clicar no botão Play no canto inferior esquerdo, que uma nova janela será aberta contendo todas as informações da análise e resultado dos cálculos da métrica, quadro a quadro do vídeo (ou vídeos, no caso de métricas com referência). É possível observar a métrica aplicada no quadro selecionado, e também um gráfico que mostra

o resultado do cálculo da métrica ao longo do vídeo, e outro abaixo com um espaço de quadros específicos que pode ser ajustado pelo usuário, assim como mostrado na Figura 7.

Figura 7 – Gráfico de saída produzidos pelo MSU VQMT



Além dos gráficos fornecidos, o MSU VQMT ainda fornece uma tabela com o resultado das métricas quadro por quadro, e algumas outras informações como a média desses resultados, o valor mínimo e máximo encontrados, quais frames possuem esses valores mínimos e máximos, a variância, entre outros dados importantes. A ferramenta também possui um campo de logs, onde é possível observar a hora que cada processo ocorreu, e demonstrar detalhadamente cada atividade realizada, como a leitura do vídeo, as informações coletadas sobre o mesmo, qual métrica foi aplicada, e se o procedimento foi realizado com sucesso ou não.

Dessa forma, com uma análise profunda e cálculos exatos é possível obter um resultado confiável da qualidade do vídeo em questão. As imagens quadro a quadro e o gráfico da métrica facilitam a compreensão do resultado obtido e tornam mais fácil a aplicação no MSU VQMT seja pra qual for sua aplicação.

Capítulo 4

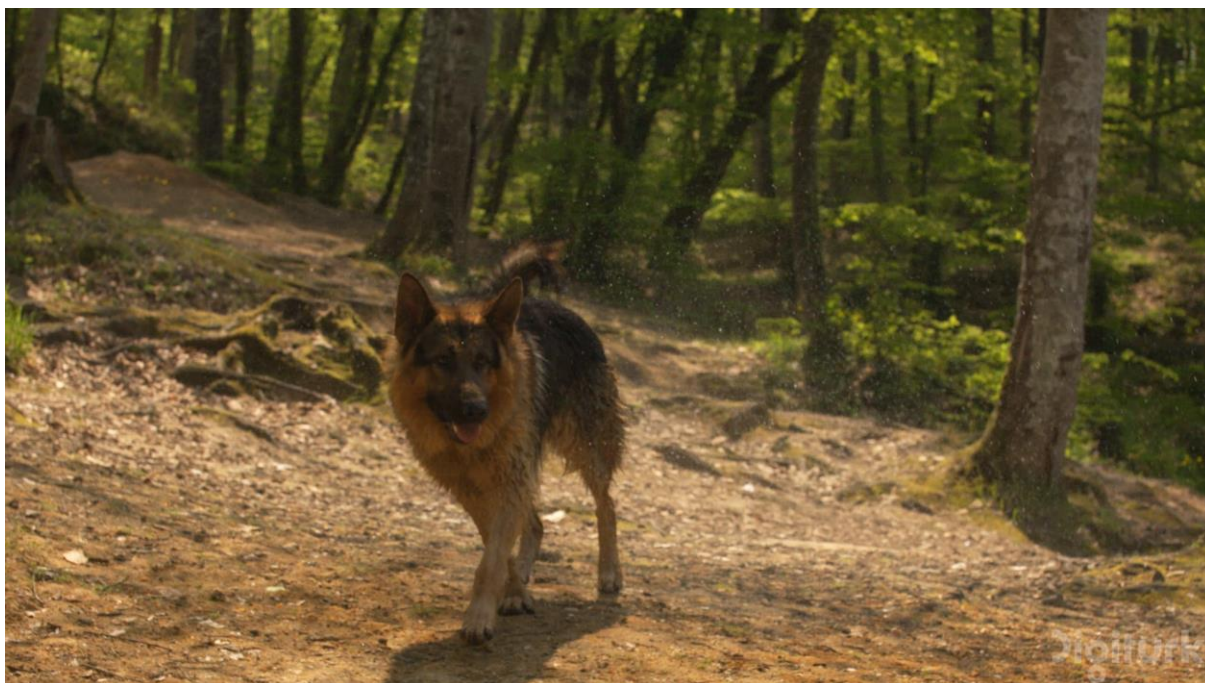
Procedimentos Experimentais e Resultados

4.1 Resultados Vídeo 1: Shake N Dry

Para o vídeo Shake N Dry, após transmitido e armazenado com o Codec H.264, foram aplicadas duas métricas com referência e duas métricas sem referência com o objetivo de obter a avaliação da qualidade do vídeo, e conseqüentemente, a QoE (Qualidade de Experiência).

O vídeo em questão possui 200 quadros. O quadro 100 foi escolhido como amostra para representar o vídeo original, o vídeo processado (transmitido) e as diferentes métricas aplicadas. O frame 100 do vídeo original pode ser observado a seguir na Figura 6. Após a transmissão do vídeo, obtemos o vídeo processado, que pode ser observado a seguir na Figura 7.

Figura 8 – Frame 100 do vídeo Shake N Dry original



Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020)

Tendo em vista que para aplicar as métricas de avaliação de qualidade com referência é necessário tanto o vídeo original, quanto o processado, foi possível o uso das mesmas. Foram selecionadas duas métricas com referência: PSNR e SSIM. O software MSU Video Quality Measurement Tool permite obter a imagem das métricas aplicadas em todos os quadros do

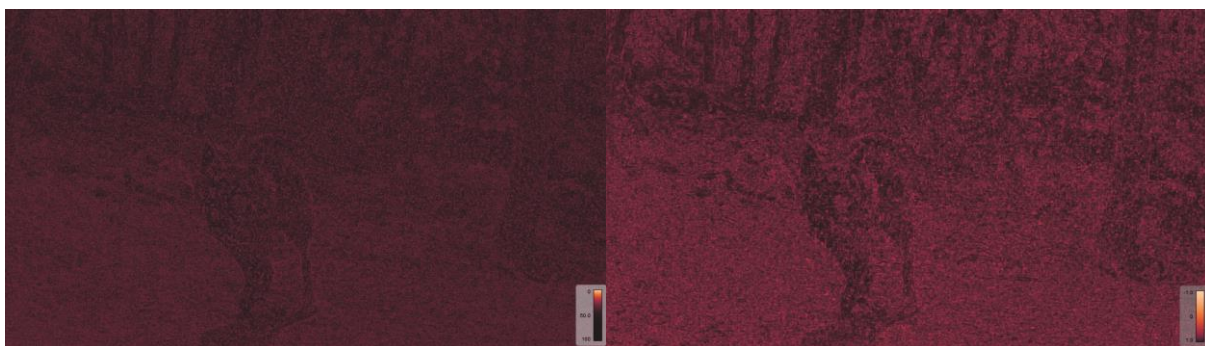
vídeo. Nas Figura 8 é possível observar o resultado das métricas aplicadas no frame 100 do vídeo Shake N Dry transmitido

Figura 9 – Frame 100 do vídeo Shake N Dry transmitido.



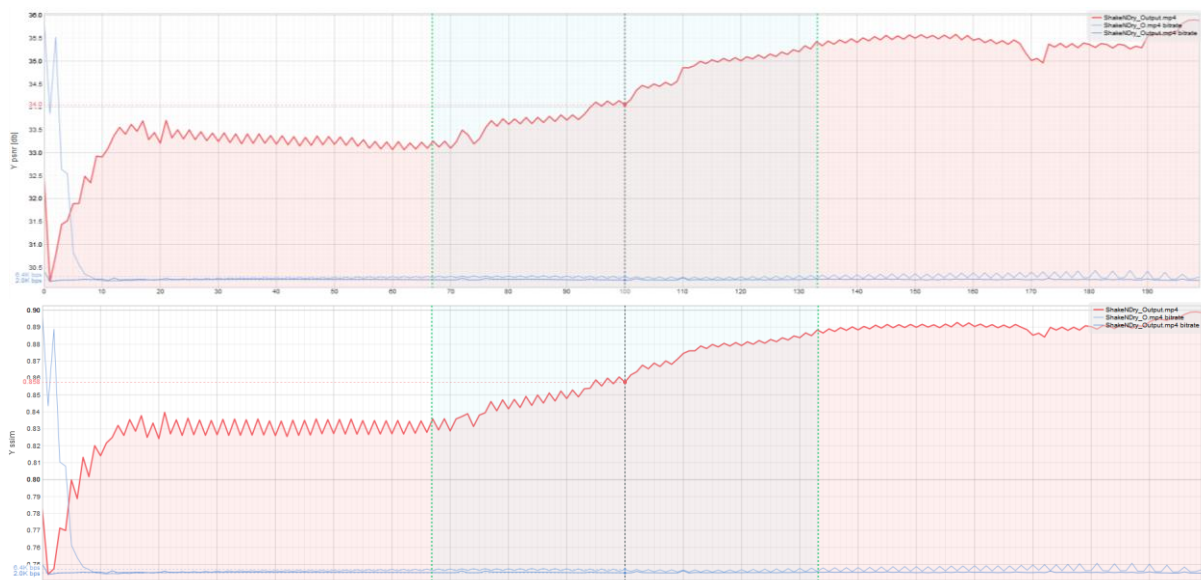
Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020) (Adaptado)

Figura 10 – À esquerda, o quadro 100 do vídeo Shak N Dry transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM



A ferramenta MSU VQMT também disponibiliza um gráfico do cálculo da métrica selecionada em todos os frames, possibilitando uma ampla visão sobre a qualidade ao decorrer de todo o vídeo. Os gráficos representam o valor resultante da métrica do eixo Y e o quadro correspondente do vídeo no eixo X. A Figura 9 mostra o gráfico para as métricas PSNR e SSIM respectivamente.

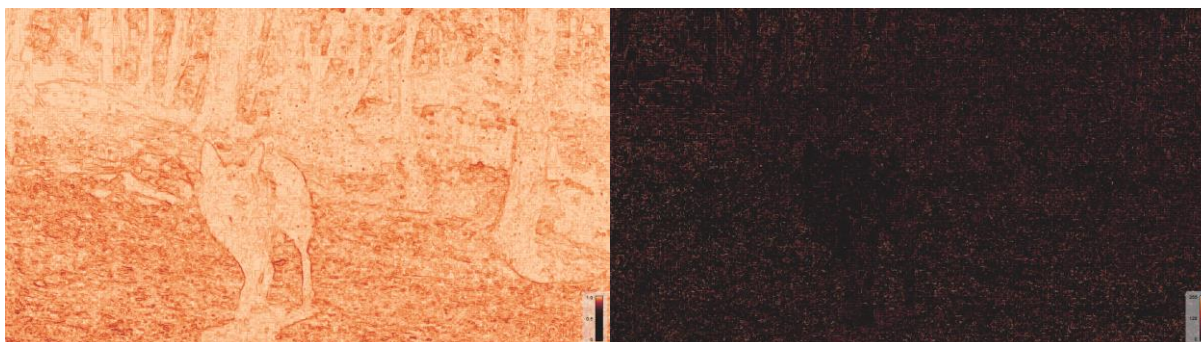
Figura 11 – Na imagem acima, o gráfico PSNR para o vídeo Shake N Dry transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Shake N Dry transmitido



A ferramenta MSU VQMT também permite a obtenção de um valor médio dos valores obtidos de todos os frames. O valor médio será tomado como base para o resultado das métricas. Para a métrica PSNR, o valor médio obtido para o vídeo ShakeNDry foi 34.258705. Para a métrica SSIM, o valor médio obtido foi 0.858383.

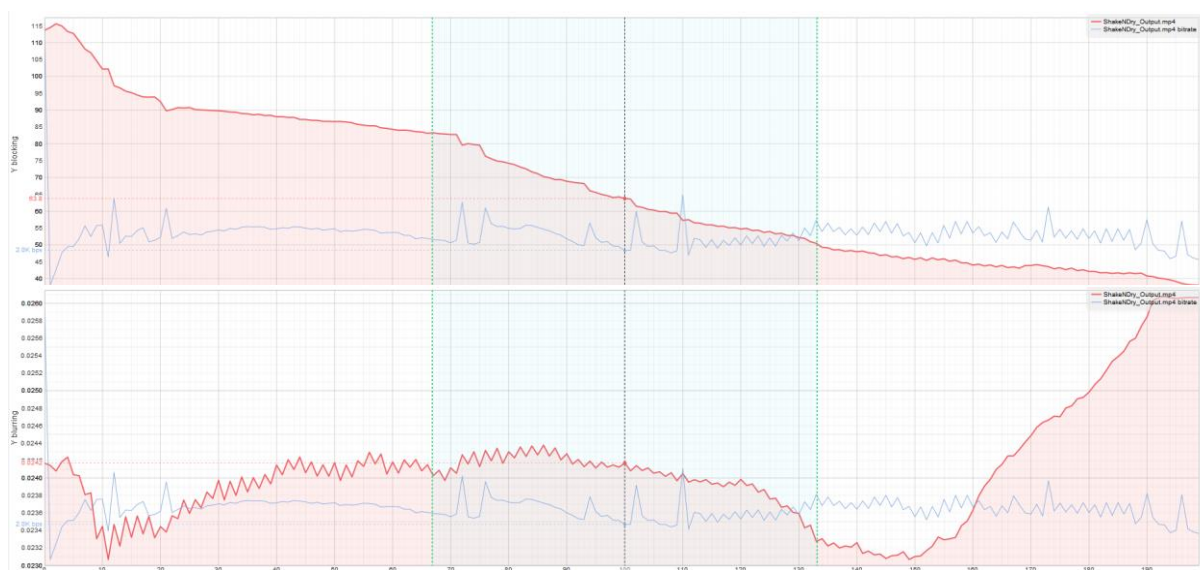
Para a avaliação da qualidade do vídeo com métricas sem referência, foram escolhidas as métricas Blurring e Blocking. Nesse tipo de métrica, o vídeo original não é necessário, sendo apenas analisado o vídeo processado. Na Figura 10 é possível observar o resultado das métricas aplicadas no quadro 100 do vídeo ShakeNDry transmitido.

Figura 12 – À esquerda, o quadro 100 do vídeo transmitida com a métrica Blocking. À direita, o mesmo quadro com a métrica Blurring.



A Figura 11, mostra os gráficos obtidos para as métricas Blocking e Blurring respectivamente. Para a métrica Blocking, o valor médio obtido para o vídeo ShakeNDry foi 67.086555. Para a métrica Blurring, o valor médio obtido foi 0.024086.

Figura 13 – Na imagem acima, o gráfico Blocking para o vídeo Shake N Dry transmitido. Na imagem abaixo, o gráfico Blurring para o vídeo Shake N Dry transmitido

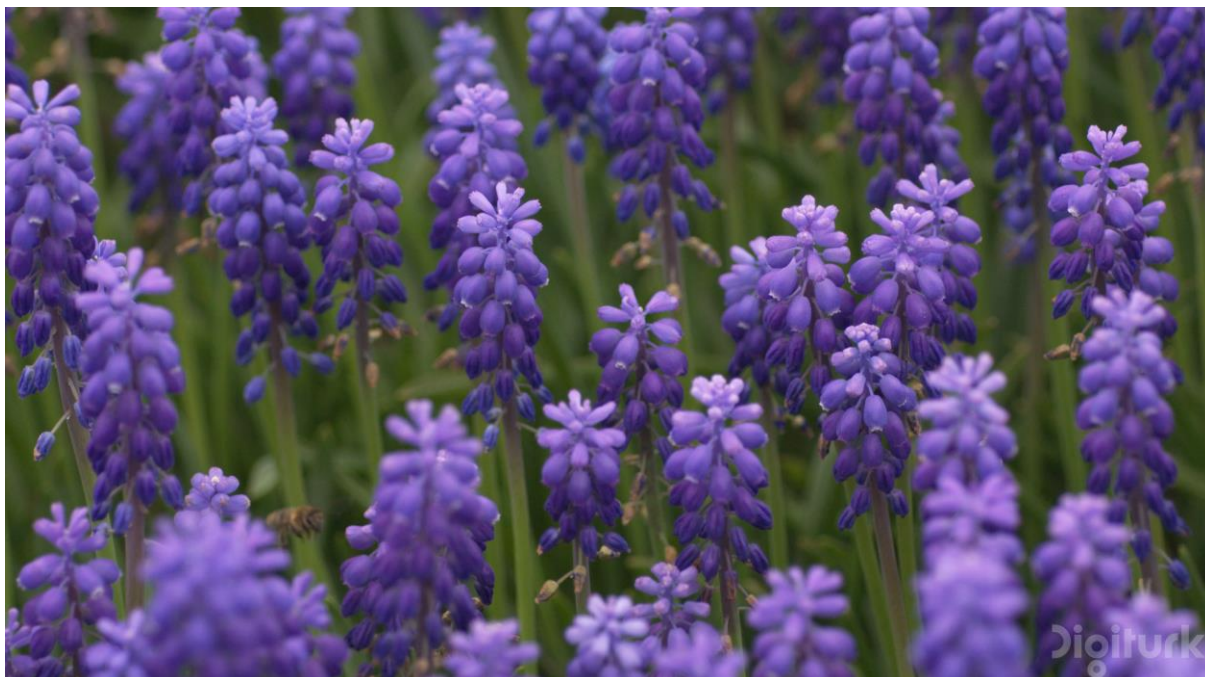


4.2 Resultados Vídeo 2: HoneyBee

Da mesma forma, para o vídeo Honey Bee, após transmitido e armazenado com o Codec H.264, foram aplicadas duas métricas com referência e duas métricas sem referência com o objetivo de obter a avaliação da qualidade do vídeo, e consequentemente, a QoE (Qualidade de experiência). O vídeo em questão possui 400 quadros. O quadro 200 foi escolhido como amostra para representar o vídeo original, o vídeo processado (transmitido) e as diferentes métricas aplicadas. O frame 200 do vídeo original pode ser observado a seguir na Figura 12. O frame 200 do vídeo transmitido pode ser observado na Figura 13.

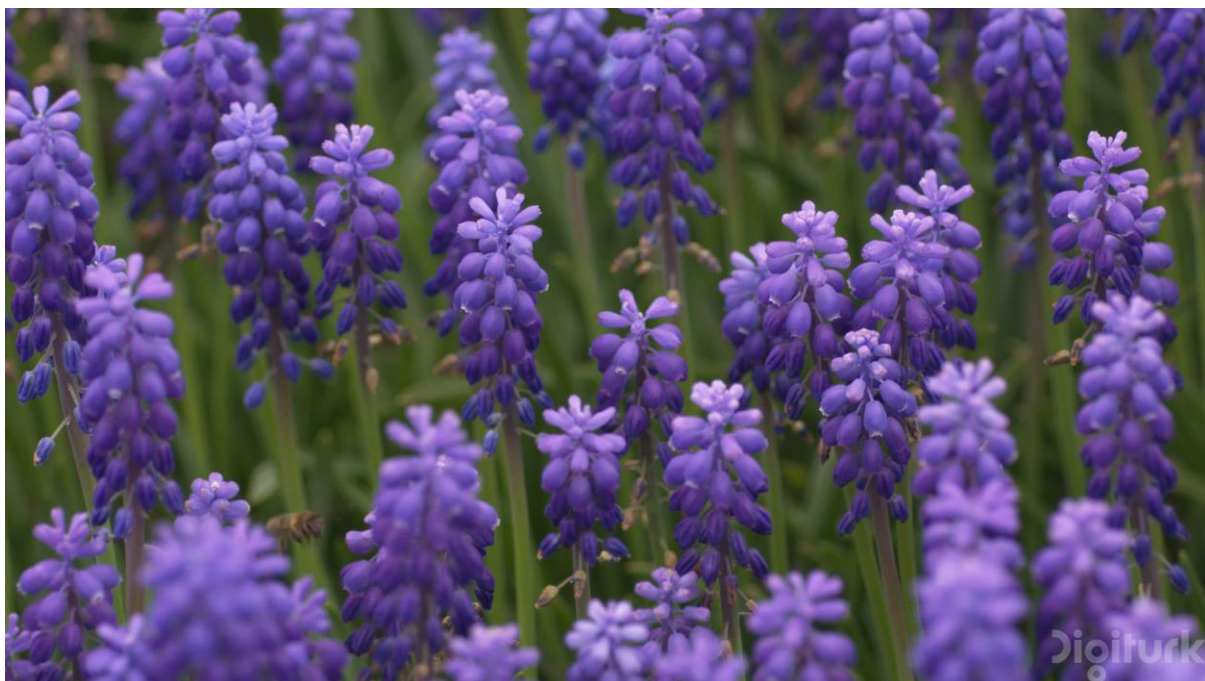
Após obtidos os vídeos original e transmitido, foram selecionadas duas métricas com referência: PSNR e SSIM. Utilizando o software MSU Video Quality Measurement Tool foi obtida a imagem das métricas aplicadas em todos os quadros do vídeo. Na Figura 14 é possível observar o resultado das métricas aplicadas no frame 200 do vídeo Honey Bee transmitido

Figura 14 – Quadro 200 do vídeo Honey Bee original



Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020)

Figura 15 – Quadro 200 do vídeo Honey Bee transmitido



Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020) (Adaptado)

A ferramenta MSU VQMT também permite a obtenção de um valor médio dos valores obtidos de todos os frames. O valor médio será tomado como base para o resultado das métricas. Para a métrica PSNR, o valor médio obtido para o vídeo Honey Bee foi 40.910431. Para a

métrica SSIM, o valor médio obtido foi 0.952582. A Figura 15, mostra os gráficos obtidos para as métricas PSNR e SSIM respectivamente.

Figura 16 – À esquerda, o quadro 200 do vídeo Honet Bee transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM

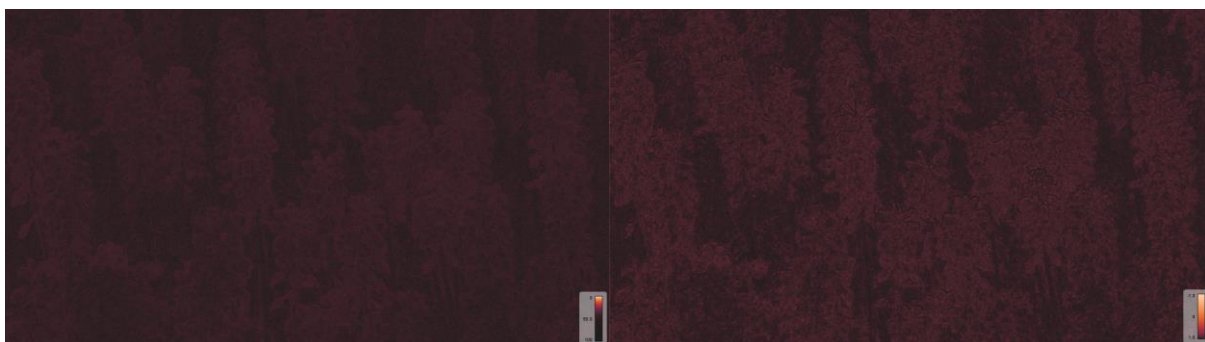


Figura 17 – Na imagem acima, o gráfico PSNR para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Honey Bee transmitido.



Como na aplicação para o vídeo She N Dry, a avaliação da qualidade do vídeo com métricas sem referência, foram escolhidas as métricas Blurring e Blocking, onde o vídeo original não é necessário, sendo apenas analisado o vídeo processado. Nas Figura 16 a seguir é possível observar o resultado das métricas aplicadas no quadro 200 do vídeo Honey Bee transmitido.

A Figura 17, mostra os gráficos obtidos para as métricas Blocking e Blurring respectivamente. Para a métrica Blocking, o valor médio obtido para o vídeo ShakeNDry foi 23.499626. Para a métrica Blurring, o valor médio obtido foi 0.015623

Figura 19 – À esquerda, o quadro 200 do vídeo Honey Bee transmitido com a métrica Blocking. À direita, o mesmo quadro com a métrica Blurring

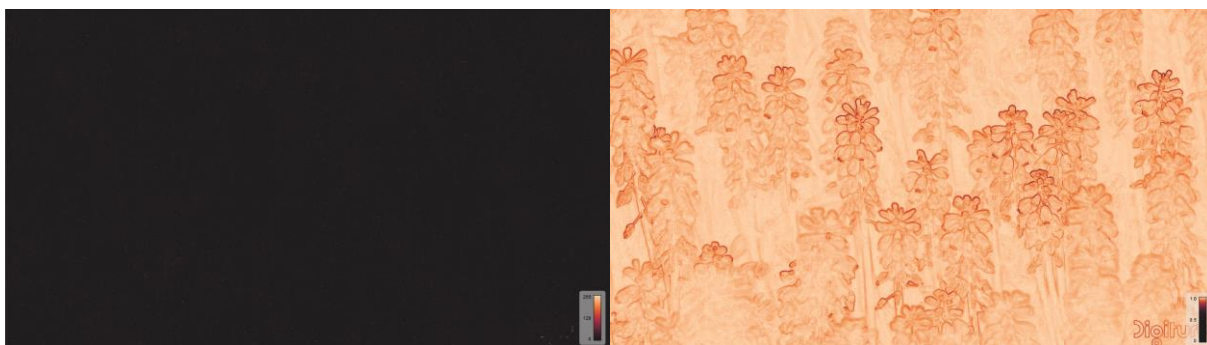
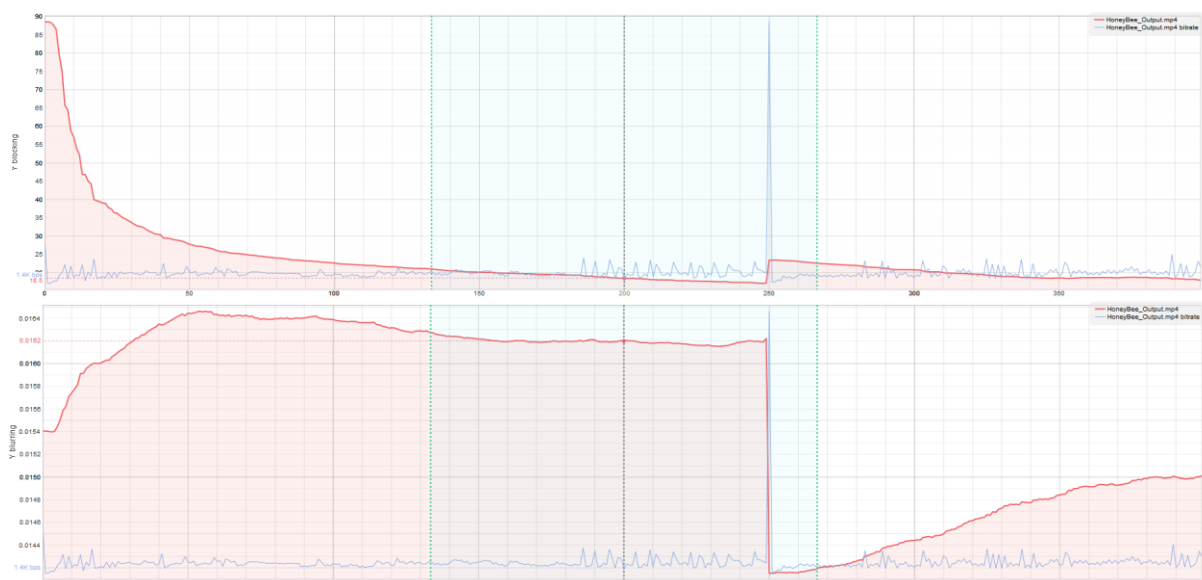


Figura 18 – Na imagem acima, o gráfico Blocking para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico Blurring para o vídeo Honey Bee transmitido



4.2 Resultados Vídeo 3: Jockey

Assim como realizado anteriormente para os vídeos Shake N Dry e Honey Bee, para o vídeo Jockey, após transmitido e armazenado com o Codec H.264, foram aplicadas duas métricas com referência e duas métricas sem referência com o objetivo de obter a avaliação da qualidade do vídeo, e conseqüentemente, a QoE (Qualidade de Experiência).

O vídeo em questão possui 400 quadros. O quadro 200 foi escolhido como amostra para representar o vídeo original, o vídeo processado (transmitido) e as diferentes métricas aplicadas.

O frame 200 do vídeo original pode ser observado a seguir na Figura 18. O frame 200 do vídeo transmitido pode ser observado na Figura 19.

Figura 20 – Quadro 200 do vídeo Jockey original.



Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020)

Figura 21 – Quadro 200 do vídeo Jockey transmitido.



Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020) (Adaptado)

Após obtidos os vídeos original e transmitido, foram selecionadas duas métricas com referência: PSNR e SSIM. Utilizando o software MSU Video Quality Measurement Tool foi obtida a imagem das métricas aplicadas em todos os quadros do vídeo. Na Figura 20 é possível observar o resultado das métricas aplicadas no frame 200 do vídeo Jockey transmitido.

Figura 23 – Na imagem acima, o gráfico PSNR para o vídeo Jockey transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Jockey transmitido

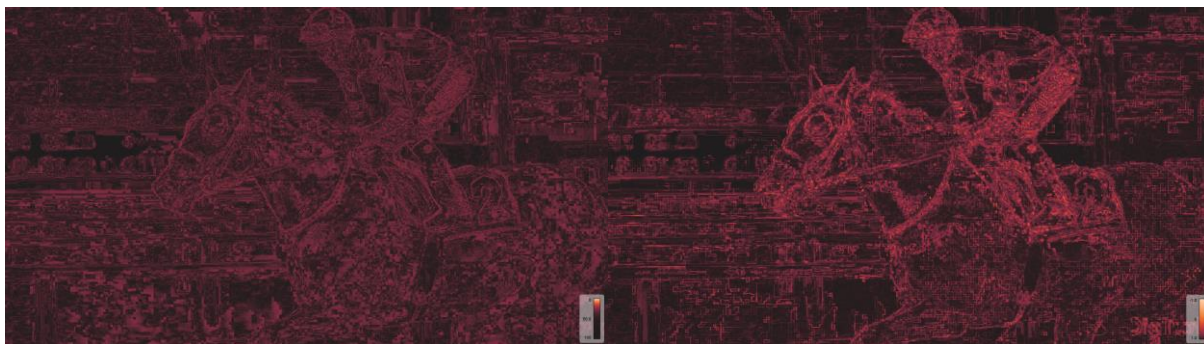
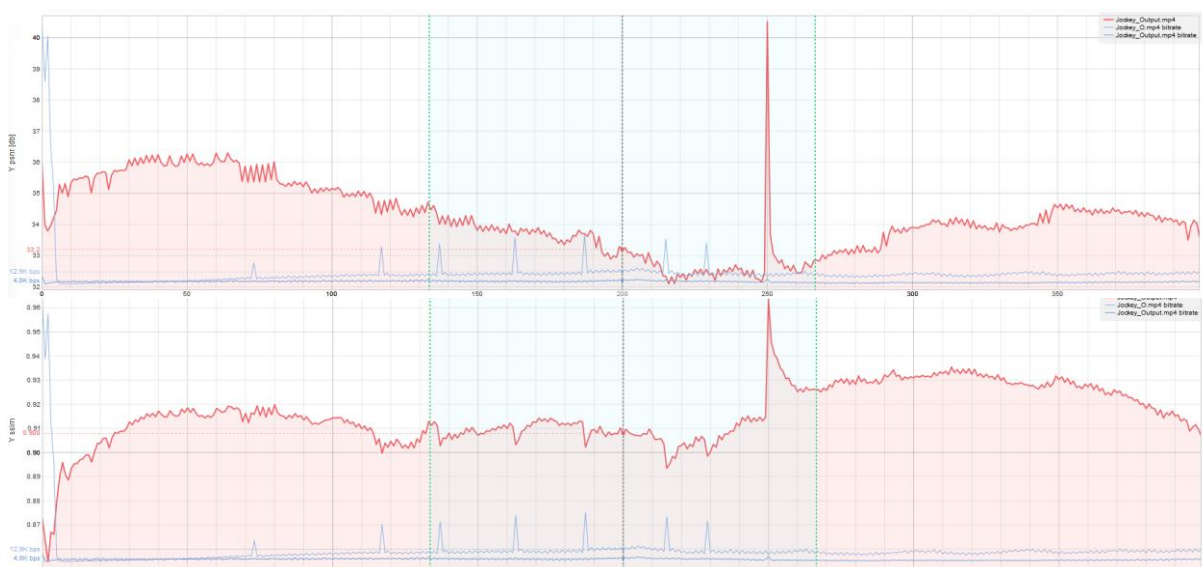


Figura 22 – À esquerda, o quadro 200 do vídeo Jockey transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM.



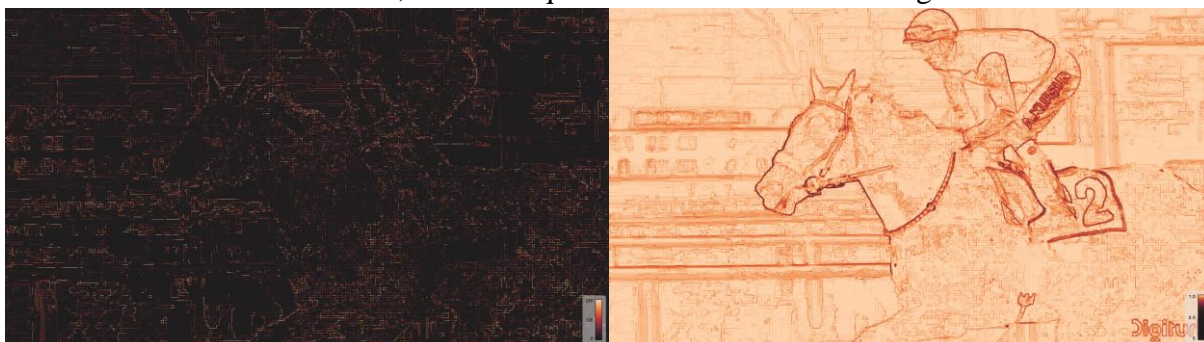
A ferramenta MSU VQMT também permite a obtenção de um valor médio dos valores obtidos de todos os frames. O valor médio será tomado como base para o resultado das métricas. Para a métrica PSNR, o valor médio obtido para o vídeo Jockey foi 34.190521. Para a métrica SSIM, o valor médio obtido foi 0.915658. A Figura 21, mostra os gráficos obtidos para as métricas PSNR e SSIM respectivamente.

Como nas aplicações anteriores, para a avaliação da qualidade do vídeo com métricas sem referência foram escolhidas as métricas Blurring e Blocking, onde o vídeo original não é

necessário, sendo apenas analisado o vídeo processado. Na Figura 22 é possível observar o resultado das métricas aplicadas no quadro 200 do vídeo Jockey transmitido.

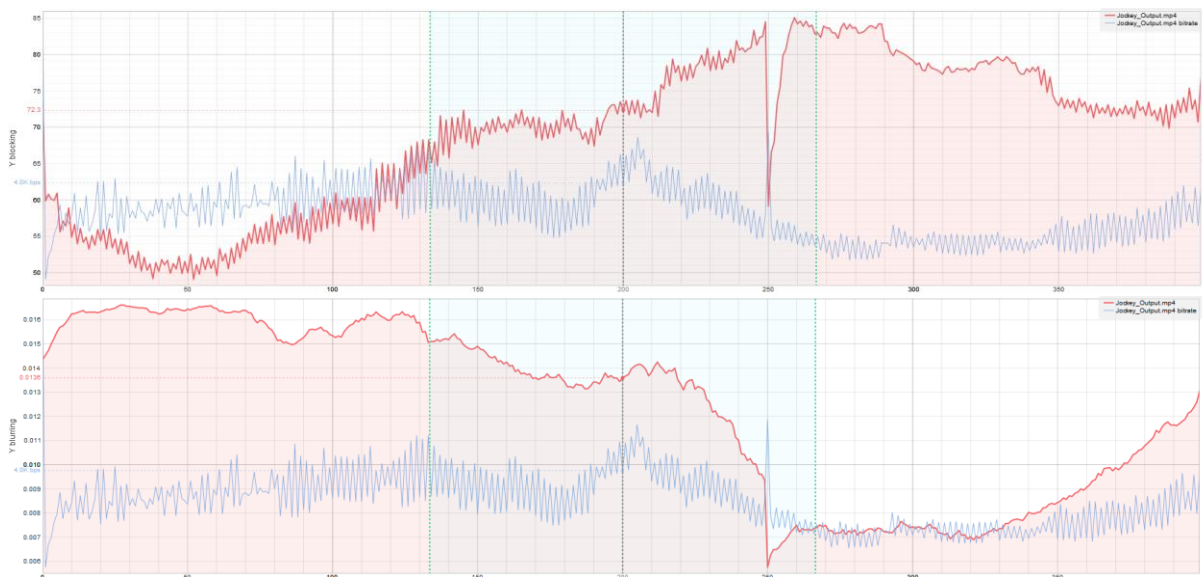
Figura 24 – À esquerda, o quadro 100 do vídeo Jockey transmitido com a métrica Blocking.

À direita, o mesmo quadro com a métrica Blurring



A Figura 23, mostra os gráficos obtidos para as métricas Blocking e Blurring respectivamente. Para a métrica Blocking, o valor médio obtido para o vídeo Jockey foi 69.048180. Para a métrica Blurring, o valor médio obtido foi 0.012391.

Figura 25 – Na imagem acima, o gráfico PSNR para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Honey Bee transmitido



4.2 Resultados Vídeo 4: Beauty

Da mesma forma como realizado anteriormente para os vídeos Shake N Dry, Honey Bee e Jockey, para o vídeo Beauty, após transmitido e armazenado com o Codec H.264, foram aplicadas duas métricas com referência e duas métricas sem referência com o objetivo de obter a avaliação da qualidade do vídeo, e consequentemente, a QoE (Qualidade de Experiência).

Figura 27 – Quadro 200 do vídeo Beauty original.



Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020)

Figura 26 – Quadro 200 do vídeo Beauty transmitido.



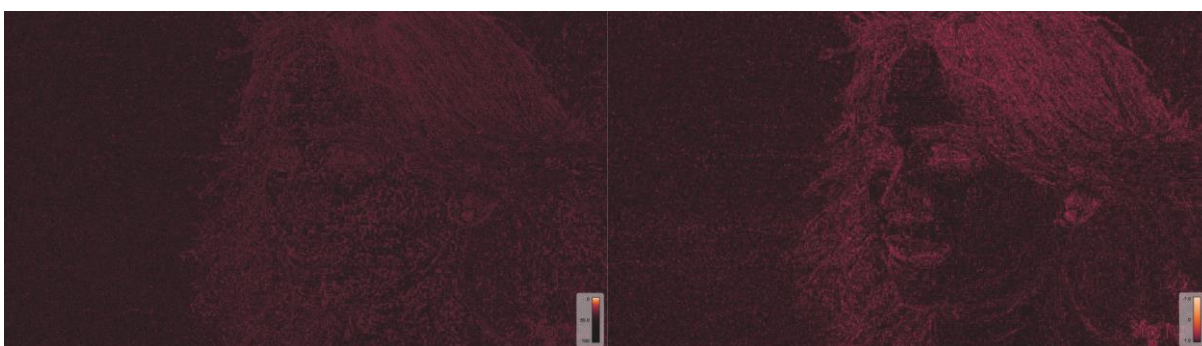
Fonte: A. Mercat, M. Viitanen, and J. Vanne. (2020) (Adaptado)

O vídeo em questão possui 400 quadros. O quadro 200 foi escolhido como amostra para representar o vídeo original, o vídeo processado (transmitido) e as diferentes métricas aplicadas.

O frame 200 do vídeo original pode ser observado a seguir na Figura 24. O frame 200 do vídeo transmitido pode ser observado na Figura 25.

Após obtidos os vídeos original e transmitido, foram selecionadas duas métricas com referência: PSNR e SSIM. Utilizando o software MSU Video Quality Measurement Tool foi obtida a imagem das métricas aplicadas em todos os quadros do vídeo. Na Figura 26 é possível observar o resultado das métricas aplicadas no frame 200 do vídeo Beauty transmitido.

Figura 28 – À esquerda, o quadro 200 do vídeo Beauty transmitido com a métrica PSNR. À direita, o mesmo quadro com a métrica SSIM

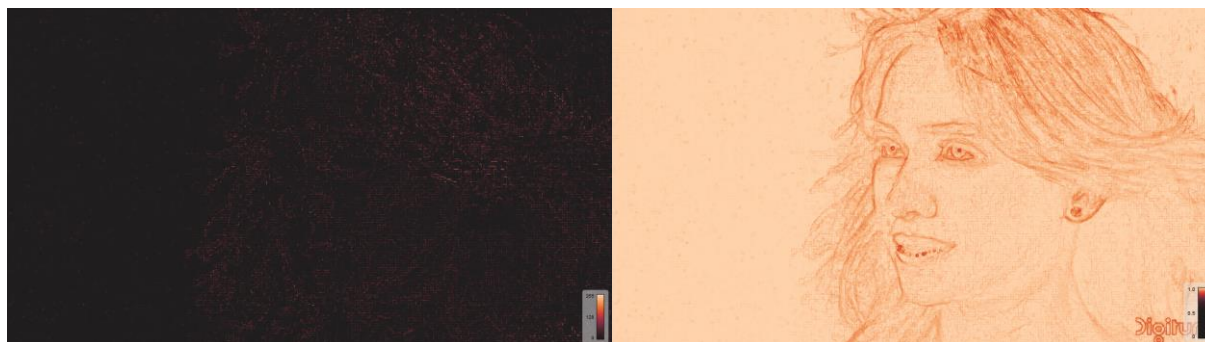


A ferramenta MSU VQMT também permite a obtenção de um valor médio dos valores obtidos de todos os frames. O valor médio será tomado como base para o resultado das métricas. Para a métrica PSNR, o valor médio obtido para o vídeo Beauty foi 38.662369. Para a métrica SSIM, o valor médio obtido foi 0.932622. A Figura 27, mostra os gráficos obtidos para as métricas PSNR e SSIM respectivamente.

Figura 29 – Na imagem acima, o gráfico PSNR para o vídeo Beauty transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Beauty transmitido.



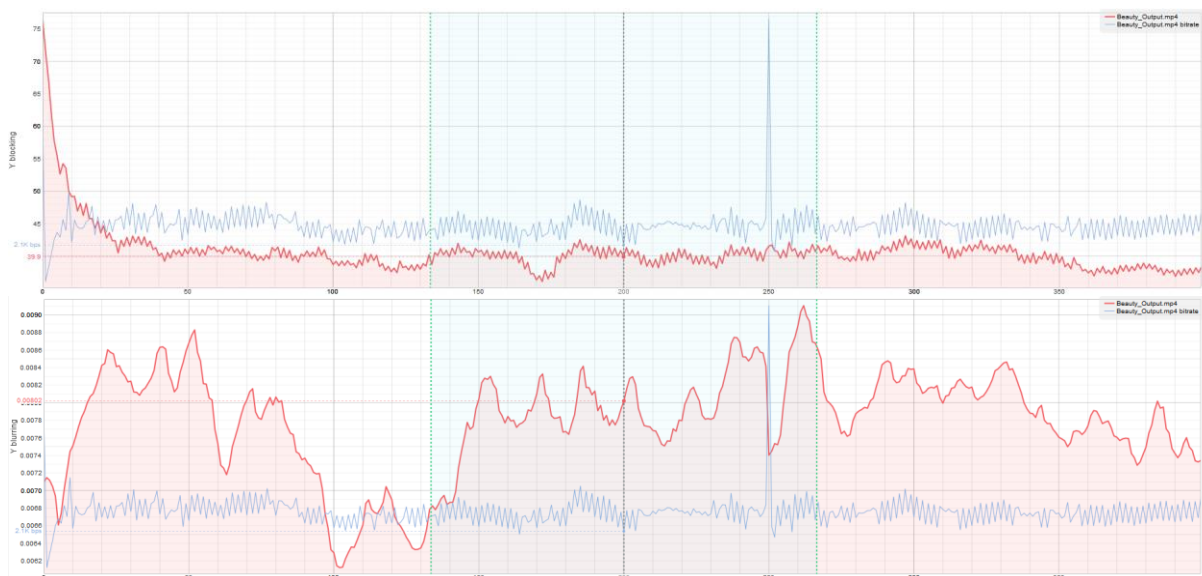
Figura 30 – À esquerda, o quadro 100 do vídeo Beauty transmitido com a métrica Blocking.
À direita, o mesmo quadro com a métrica Blurring.



Como nas aplicações anteriores, para a avaliação da qualidade do vídeo com métricas sem referência foram escolhidas as métricas Blurring e Blocking, onde o vídeo original não é necessário, sendo apenas analisado o vídeo processado. Na Figura 28 é possível observar o resultado das métricas aplicadas no quadro 200 do vídeo Jockey transmitido.

A Figura 29, mostra os gráficos obtidos para as métricas Blocking e Blurring respectivamente. Para a métrica Blocking, o valor médio obtido para o vídeo Beauty foi 40.682709. Para a métrica Blurring, o valor médio obtido foi 0.007807.

Figura 31 – Na imagem acima, o gráfico PSNR para o vídeo Honey Bee transmitido. Na imagem abaixo, o gráfico SSIM para o vídeo Honey Bee transmitido.



4.5 Análise dos Resultados

Com a conclusão dos procedimentos experimentais, é possível organizar os resultados obtidos de uma forma que facilite o entendimento dos mesmos para cada um dos vídeos e

métricas utilizados. O formato de tabelas foi considerado o mais adequado para a organização e visualização dos resultados.

Para uma melhor compreensão dos resultados, as Tabelas 3, 4, 5 e 6 mostram alguns valores utilizados como referência para as métricas PSNR, SSIM, Blocking e Blurring respectivamente. Esses valores de referência servirão como base para a análise dos resultados das médias obtidas no procedimento experimental.

Tabela 3 - Valores de referência para a métrica PSNR.

Classificação	Valor PSNR (dB)
Excelente	0
Baixo	100

Tabela 4 - Valores de referência para a métrica SSIM.

Classificação	Valor SSIM
Excelente	1
Baixo	-1

Tabela 5 - Valores de referência para a métrica Blocking.

Classificação	Valor Blocking
Excelente	0
Baixo	∞

Tabela 6 - Valores de referência para a métrica Blurring.

Classificação	Valor Blurring
Excelente	0
Baixo	1

Com os valores de referência estabelecidos, é possível realizar a análise dos resultados das médias obtidas para as métricas aplicadas nos vídeos. A Tabela 7 mostra os valores obtidos para o vídeo ShakeNDry. Para a métrica PSNR, foi obtido o valor médio de 34,258705, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica SSIM, foi obtido o valor médio de 0,858383, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica Blocking, foi obtido o valor médio de 67,086555, que de acordo com os valores de referência é considerado um resultado aceitável. Para a métrica Blurring, foi obtido o valor médio de 0,024086, que de acordo com os valores de referência é considerado um resultado excelente.

Tabela 7 - Valores obtidos para o vídeo ShakeNDry

Métrica	Valor médio
PSNR	34,258705
SSIM	0,858383
Blocking	67,086555
Blurring	0,024086

A Tabela 8 mostra os valores obtidos para o vídeo HoneyBee. Para a métrica PSNR, foi obtido o valor médio de 40,910431, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica SSIM, foi obtido o valor médio de 0,952582, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica Blocking, foi obtido o valor médio de 23,499626, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica Blurring, foi obtido o valor médio de 0,015623, que de acordo com os valores de referência é considerado um resultado excelente.

Tabela 8 - Valores obtidos para o vídeo HoneyBee

Métrica	Valor médio
PSNR	40,910431
SSIM	0,952582
Blocking	23,499626
Blurring	0,015623

A Tabela 9 mostra os valores obtidos para o vídeo Jockey. Para a métrica PSNR, foi obtido o valor médio de 34,190521, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica SSIM, foi obtido o valor médio de 0,915658, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica Blocking, foi obtido o valor médio de 69,048180, que de acordo com os valores de referência é considerado um resultado aceitável. Para a métrica Blurring, foi obtido o valor médio de 0,012391, que de acordo com os valores de referência é considerado um resultado excelente.

Tabela 9 - Valores obtidos para o vídeo Jockey.

Métrica	Valor médio
PSNR	34,190521
SSIM	0,915658
Blocking	69,048180
Blurring	0,012391

A Tabela10 mostra os valores obtidos para o vídeo Beauty. Para a métrica PSNR, foi obtido o valor médio de 38,662369, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica SSIM, foi obtido o valor médio de 0,932622, que de

acordo com os valores de referência é considerado um resultado excelente. Para a métrica Blocking, foi obtido o valor médio de 40,682709, que de acordo com os valores de referência é considerado um resultado excelente. Para a métrica Blurring, foi obtido o valor médio de 0,007807, que de acordo com os valores de referência é considerado um resultado excelente.

Tabela 10 - Valores obtidos para o vídeo Beauty.

Métrica	Valor médio
PSNR	38,662369
SSIM	0,932622
Blocking	40,682709
Blurring	0,007807

Ao término da análise dos resultados, pode-se concluir que os resultados obtidos pelo valor médio das métricas aplicadas foi satisfatório e se mostrou excelente na maioria das vezes, afirmando que a qualidade de experiência (QoE) em vídeos transmitidos com o codec H.264 é ótimo. Os resultados que se mostraram menos aceitáveis foram os referentes a métrica Blocking. Isso se deve pelo grande efeito do artefato blocking exercido durante a transmissão de vídeo. O blocking se dá principalmente na utilização de codecs com perdas, onde o processo de compressão pode causar danos ao vídeo. Como o H.264 é um codec com perdas, é compreensível que este artefato seja encontrado em algumas das amostras utilizadas.

Capítulo 5

Conclusão

5.1 Considerações Finais

Neste trabalho foi realizada a avaliação da qualidade de experiência em transmissão de vídeo utilizando o codec H.264, utilizando duas métricas de avaliação de qualidade de vídeo com referência completa e duas métricas sem referência.

As quatro amostras de vídeos utilizadas para realizar a transmissão e a avaliação foram obtidas através do conjunto de dados abertos do Ultra Vídeo Group (UVG), que fornece versáteis sequências de vídeos em 4K com diversas características, onde o usuário pode escolher a que melhor se adequa ao seu projeto. As sequencias escolhidas foram: ShakeNDry, um vídeo que exibe um cão caminhando por uma floresta; HoneyBee, um vídeo que exibe uma abelha voando entre várias flores; Jockey, um vídeo que exibe um cavalo em uma pista de corrida; e Beauty, um vídeo que exibe o rosto de uma mulher em um fundo preto.

Após as amostras obtidas, foi necessária realizar a transmissão com o codec H.264. Para isso, foi utilizada a ferramenta FFmpeg, um sistema que permite gravar, converter e transmitir áudio e vídeos em diversos formatos. Foi necessário o uso de dois computadores e através de uma linha de comando aplicado do prompt de comando do sistema operacional Windows, foi realizada a transmissão utilizando o protocolo UDP, um protocolo de comunicação amplamente utilizado em transmissões de vídeos.

Com os vídeos transmitidos com o codec H.264 através do protocolo UDP, foi possível aplicar as métricas de avaliação de qualidade objetiva. Foram escolhidas quatro métricas, duas de referência completa (PSNR e SSIM) e duas sem referência (Blocking e Blurring). As métricas de referência completa necessitam do vídeo original e do processado, nesse caso o transmitido, para realizar um processo de comparação entre algumas características de ambos os arquivos. As métricas sem referência não necessitam do vídeo original, sendo avaliado somente o vídeo processado com base em alguns aspectos.

Para a realização da avaliação de qualidade dos vídeos transmitidos, foi utilizado o software MSU Video Quality Measure Tool, uma ferramenta de análise profunda de vídeo. Ela permite a aplicação de várias métricas de VQA, entre elas as quatro escolhidas para esse

projeto, e fornece imagens e gráficos quadro a quadro dos resultados obtidos, além de vários outros artefatos, como a média dos valores obtidos para todos os quadros do vídeo. Uma vez as métricas aplicadas para todos os vídeos, foi utilizado o valor médio das métricas para os vídeos. Após os valores médios obtidos, foi realizada uma análise dos resultados, que se mostraram bastante satisfatórios.

Conclui-se que a maioria dos resultados obtidos das métricas de avaliação de qualidade para os vídeos transmitidos com o codec H.264 foram excelentes. Somente alguns resultados se demonstraram satisfatórios para a métrica blocking devido a presença do artefato efeito de blocagem, que se mostrou presente em dois dos vídeos transmitidos. Como o H.264 é um codec com perdas, é compreensível que o artefato blocking esteja presente em algumas amostras, como resultado do processo de compressão ocorrido. Dessa forma, a qualidade de experiência dos usuários do codec H.264 se demonstrou satisfatória, baseada na aplicação das métricas de avaliação de qualidade.

5.2 Trabalhos Futuros

- Avaliar a qualidade de experiência em transmissão de vídeo utilizando outros codecs, como o H.265 por exemplo;
- Realizar a transmissão de vídeo utilizando outros protocolos, como o TCP ou HTML por exemplo;
- Aplicação de outras métricas de referência completa, referência reduzida e sem referência nos vídeos transmitidos;
- Realizar a transmissão do vídeo em diferentes ambientes e com diferentes distâncias entre o transmissor e emissor para avaliar a influência dessas características na qualidade do vídeo;

Referências

NIDHI, & AGGARWAL, N. A review on Video Quality Assessment. Recent Advances in Engineering and Computational Sciences (RAECS). 2014

ROSSHOLM, Andreas; SHAHID, Muhammad; LÖVSTRÖM, Benny. Comparison of machine learning methods for quality estimation of videos with diversity in temporal, spatial, and quantization domain. Submitted to a journal, Nov 2014.

KORBEL, Frantisek. FFmpeg Basics Multimedia handling with a fast audio and video encode. 2012.

AKRAMULLAH, Shahriar. Digital Video Concepts, Methods, and Metrics. A press open Springer, 2014.

ZHU, Kongfeng. No-reference Video Quality Assessment and Applications. 2014.

PANDREMMENOU, Katerina; SHAHID, Muhammad; KONDI, Lisimachos P.; LÖVSTRÖM, Benny. A No-Reference Bitstream-based Perceptual Model for Video Quality Estimation of Videos Affected by Coding Artifacts and Packet Losses. 2014.

NEHMÉ, Yanal; FARRUGIA; Jean-Philippe; DUPONT, Florent; LE CALLET, Patrick; LAVOUÉ, Guillaume. Comparison of Subjective Methods for Quality Assessment of 3D Graphics in Virtual Reality. ACM Transactions on Applied Perception, v. 37, n. 4, Article 111, August 2020.

FARIAS, Mylene C.Q. Video Quality Metrics. Universidade de Brasília, 2010.

POOBATHY, D.; CHEZIAN, Manicka. Edge Detection Operators: Peak Signal to Noise Ratio Based Comparison. 2014.

SUN, Huifang; VETRO, Anthony; XI, Jun. An overview of scalable video streaming. 2007.

DAVCEV, D.; GÓMEZ, J. M. (Eds.). ICT Innovations 2009. 2010.

NEWMARCH, Jan. Linux Sound Programming. 2017.

WINKLER, Stefan. Digital Video Quality: vision models and metrics. 2005.

SHAHID, M; ROSSHOLM, Andreas; LÖVSTRÖM, Benny; ZUPERNICK, Hans-Jürgen. No-reference image and video quality assessment: a classification and review of recent approaches. EURASIP Journal on Image and Video Processing, 2014.

VQMT Metrics. Disponível em: <<https://videoprocessing.ai/vqmt/metrics/>>. Acesso em: 15 de outubro de 2023.

WU, HR; RAO, KR. Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications). CRC, Boca Raton, 2005.

POSTEL, J. RFC 768—User Datagram Protocol. August 1980.

RICHARDSON, Iain E. The H.264 Advanced Video Compression Standard. Second Edition. 2010.

CHEN, Jian-Wen; KAO, Chao-Yang; LIN, Youn-Long. Introduction to H.264 Advanced Video Coding.

MSU VQMT Brief Description. Disponível em: <<https://videoprocessing.ai/vqmt/basic/>>. Acesso em: 25 de outubro de 2023.

APOSTOLOPOULOS, John G.; TAN, Waitian; WEE, Susie J. Video Streaming: concepts, algorithms and systems. 2002.

MERCAT, A.; VIITANEN, M.; VANNE, J. UVG dataset: 50/120fps 4K sequences for video codec analysis and development. In: ACM Multimedia System Conference, 2020, Istanbul, Turkey.