



**UNIVERSIDADE FEDERAL DO PARÁ
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

ANDRÉ FIGUEIRA RIKER

**UM MECANISMO DE ESTIMATIVA DE
QUALIDADE DE EXPERIÊNCIA EM TEMPO
REAL PARA SISTEMAS MULTIMÍDIA: ESTUDO
DE CASO PARA REDES EM MALHA SEM FIO**

BELÉM-PA

Mar / 2012

ANDRÉ FIGUEIRA RIKER

**UM MECANISMO DE ESTIMATIVA DE QUALIDADE DE
EXPERIÊNCIA EM TEMPO REAL PARA SISTEMAS
MULTIMÍDIA: ESTUDO DE CASO PARA REDES EM
MALHA SEM FIO**

Dissertação submetida à banca julgadora na
Universidade Federal do Pará como parte dos
requisitos para obtenção do grau de Mestre em
Ciência da Computação

Orientador: Dr. Antônio Jorge Gomes Abelém

Co-Orientador: Dr. Eduardo Coelho Cerqueira

BELÉM-PA

Mar / 2012

ANDRÉ FIGUEIRA RIKER

**UM MECANISMO DE ESTIMATIVA DE
QUALIDADE DE EXPERIÊNCIA EM TEMPO
REAL PARA SISTEMAS MULTIMÍDIA: ESTUDO
DE CASO PARA REDES EM MALHA SEM FIO**

Dissertação submetida à banca julgadora na
Universidade Federal do Pará como parte dos
requisitos para obtenção do grau de Mestre
em Ciência da Computação

Aprovada em: --/--/----

BANCA EXAMINADORA

Prof. Dr. Antônio Jorge Gomes Abelém
Universidade Federal do Pará
Orientador

Prof. Dr. Eduardo Coelho Cerqueira
Universidade Federal do Pará
Co-orientador

Prof. Dr. Luciano Paschoal Gaspar
Universidade Federal do Rio Grande do Sul

Prof. Dr. Augusto José Venâncio Neto
Universidade Federal do Ceará

Aos meus pais.

Agradecimentos

Ao contrário do que está estampado na capa, esta dissertação de mestrado não pertence somente a mim. Digo isto com convicção porque eu e Deus sabemos que pessoas me ajudaram muito nesta batalha. Por isso eu partilho este título com muitos. Alguns deles, não são mestres em nenhuma área da ciência, mas são verdadeiros mestres da vida. Por outro lado, alguns foram professores que serviram de grande inspiração.

Hoje, olhando para trás, lembro bem que houveram momentos difíceis. Nesses momentos, não foram as fórmulas, nem as equações que me ajudaram. Porém foram pessoas que me fizeram avançar. Por isso, agradeço à maior delas, Jesus Cristo, o mestre dos mestres, o qual nunca estudou em uma universidade mas mudou o mundo com seu amor.

Não poderia deixar de agradecer aos meus pais. Reconheço que na visão deles, completar este curso não muda muita coisa. Na verdade, sei que o amor deles não é alterado pelos diplomas que recebo. Pecebo no fundo dos seus corações, que o importante é eu estar com vida e nada mais. Isso é um grande ensinamento para minha vida. Agradeço à Deus pela família que tenho. A família de três irmãos (Daniel, David e André) e do casal Delmas e Márcia. Lembrando sempre de duas pessoas especiais: Francisco Dias e Pabla Nunhes.

Além disso, fica aqui um agradecimento para os amigos: Helder e Larissa Guimarães, Jair Vaz, Judson Brito, Joe, Jana, Luis e Jane Riker, Allan Lacerda, Diana Lopes, Jully Luz e muitos outros.

Para fechar com chave de ouro. Agradeço à Andressa pelo apoio e companheirismo. Em todos os momentos, sempre me ajudou e incentivou. Essa conquista não teria graça se você não estivesse ao meu lado.

Chego ao fim de uma etapa e descubro que a vida não pára. Percebo que prosseguir não é uma opção. Por isso, devo rapidamente esquecer das coisas que para trás ficam e seguir para o alvo. Com fé em Deus chegarei lá.

Resumo

Devido ao grande crescimento das aplicações de vídeo e a implantação das redes em malha sem fio, a necessidade de mecanismos que estimem a qualidade de experiência (QoE) das aplicações de vídeo em tempo real tem se tornado essencial, uma vez que as operadoras querem controlar e otimizar seus recursos, enquanto matém os usuários de aplicações multimídia com um bom nível de qualidade. Por outro lado, do ponto de vista dos protocolos de rede, é de grande importância estar ciente do nível de qualidade percebido pelos usuários das aplicações de vídeo. Assim as políticas de gerenciamento dos recursos de rede poderiam ser aplicadas levando em consideração a percepção dos usuários. Ao contrário do que ocorre atualmente, onde o gerenciamento dos recursos é feito com base nos níveis de QoS (*Quality of Service*). Portanto, este trabalho propõe um mecanismo chamado MultiQoE que utiliza parâmetros de *codec* e rede para estimar em tempo real a qualidade de vídeos que possuam diferentes codificações, níveis de atividade espaço-temporal e degradações. Para isso, transmissões de vídeo foram realizadas em uma rede *mesh* no NS-2.34 e avaliações subjetivas foram feitas com 55 avaliadores humanos. Os resultados apresentam a alta correlação do MultiQoE com a métrica subjetiva *Mean Opinion Score* (MOS) e deixam claro a baixa precisão das métricas objetivas (*Video Quality Metric* - VQM, *Structural Similarity Index Metric* - SSIM e *Peak Signal-to-Noise Ratio* - PSNR) e do método híbrido chamado Pseudo-Subjective Quality Assessment (PSQA).

Abstract

Due to the large growth of video applications and the deployment of wireless mesh networks, the need for mechanisms to estimate the quality of experience (QoE) for video applications in real time has become essential, since the operators want to control and optimize their resources, while maintains the users of multimedia applications with a good level of quality. Moreover, from the viewpoint of network protocols, is very important to be aware of the quality level perceived by the video applications users. Thus the management policies of network resources could be applied taking into account the users perception. Contrary to what currently occurs, where resource management is based on levels of Quality of Service (QoS). Therefore, this work proposes a mechanism called MultiQoE that uses network and codec parameters to estimate in real-time the quality of videos that have different encodings, levels of spatio-temporal activity and degradation. For this, video streams were performed on a mesh network (NS 2.34) and subjective evaluations were done with 55 human evaluators. The results show a high correlation with the subjective score (MOS) and highlight the lower precision of objective metrics (Video Quality Metric - VQM, Structural Similarity Index Metric - SSIM and Peak Signal-to-Noise Ratio - PSNR) and the hybrid method called Pseudo-Subjective Quality Assessment (PSQA).

Sumário

1	Introdução	p. 2
1.1	Qualidade de experiência em redes em malha sem fio	p. 2
1.2	Objetivos	p. 3
1.3	Organização do texto	p. 4
2	Compressão de vídeo para sistemas multimídia	p. 5
2.1	Introdução à compressão de vídeo	p. 5
2.2	Princípios de compressão de vídeo	p. 7
2.2.1	Compressão temporal	p. 8
2.2.2	Compressão espacial	p. 10
2.3	Estrutura do MPEG-4	p. 12
2.4	Conclusão do Capítulo	p. 13
3	Qualidade de Experiência	p. 14
3.1	Introdução à Qualidade de Experiência	p. 14
3.2	Métodos e métricas de estimativa de QoE	p. 15
3.2.1	Métodos Subjetivos	p. 16
3.2.1.1	Single Stimulus (SS)	p. 16
3.2.1.2	Double Stimulus Impairment Scale (DSIS)	p. 18
3.2.2	Métricas Objetivas	p. 18

3.2.2.1	Peek Signal to Noise Ratio (PSNR).....	p. 19
3.2.2.2	Structural Similarity Index Metric (SSIM).....	p. 20
3.2.2.3	Video Quality Metric (VQM).....	p. 21
3.2.3	Métodos Híbridos.....	p. 22
3.3	Conclusão do Capítulo.....	p. 22
4	Trabalhos relacionados.....	p. 23
4.1	PSQA.....	p. 23
4.1.1	Implementações do PSQA.....	p. 24
4.1.1.1	Implementações em redes P2P.....	p. 25
4.1.1.2	Implementações em redes IEEE 802.11.....	p. 25
4.1.1.3	Utilização de múltiplos caminhos para transmissão de vídeo.....	p. 26
4.1.1.4	Utilização de técnicas de SVC para transmissão de vídeo.....	p. 27
4.1.2	Desvantagens do PSQA.....	p. 28
4.2	<i>Discrete Evaluation Quality Assessment</i> (DEQA).....	p. 28
4.3	MintMOS.....	p. 29
4.4	<i>Perceived QoS</i> (PQoS).....	p. 30
4.5	Classificação de conteúdo de vídeo.....	p. 31
4.6	Conclusão do Capítulo.....	p. 32
5	MultiQoE.....	p. 33
5.1	Visão geral.....	p. 33
5.2	Funcionamento.....	p. 34
5.3	Implementação.....	p. 36
5.3.1	Base de vídeos originais.....	p. 36
5.3.2	Classificação de conteúdo.....	p. 37
5.3.3	Condições de rede e composição da base de vídeos degradada.....	p. 39
5.3.4	Avaliação subjetiva.....	p. 40
5.3.5	Seleção de fatores que afetam a qualidade.....	p. 42
5.3.5.1	Taxa de perda de <i>frames</i> I, P e B.....	p. 43
5.3.5.2	Nível de atividade espaço-temporal.....	p. 43

5.3.5.3	Tamanho do GoP	p. 44
5.3.6	Múltiplas redes neurais artificiais	p. 45
5.4	Conclusão do Capítulo	p. 46
6	Estudo de caso	p. 47
6.1	Configurações do ambiente utilizado.....	p. 47
6.2	Análise dos resultados	p. 49
6.3	Conclusão do capítulo	p. 52
7	Conclusões e trabalhos futuros.....	p. 53
7.1	Conclusões gerais	p. 53
7.2	Trabalhos futuros	p. 54
7.3	Publicações	p. 55
	Referências	p. 56

Lista de abreviaturas e siglas

AEMT	Alto nível Espacial e Médio nível Temporal
ANN	Artificial Neural Network
AP	Access Point
BET	Baixo nível Espaço-Temporal
DCT	Discrete Cosine Transform
DEQA	Discrete Evaluation Quality Assessment
DSIS	Double Stimulus Impairment Scale
FEC	Forward Error Correction
FR	Full Reference
GoP	Group of Picture
GW	Gentle Walking
IDR	Instantaneous Decoder Refresh
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ITS	Institute for Telecommunication Science
ITU	International Telecommunication Union
LM	Levenberg-Marquardt
MANNs	Multiple Artificial Neural Networks
MEAT	Médio nível Espacial Alto nível Temporal
MPEG	Motion Picture Expert Group
MSE	Mean Squared Error

NR	No Reference
NS	Network Simulator
P2P	Peer-to-Peer
PSNR	Peak Signal to Noise Ratio
PSQA	Pseudo Subjective Quality Assessment
QoE	Quality of Experience
QoS	Quality of Service
RM	Rapid Movement
RNN	Random Neural Network
RR	Reduced Reference
SAD	Sum of Absolute Difference
SM	Slight Movement
SS	Single Stimulus
SSIM	Structural Similarity Index Metric
SVC	Scalable Video Coding
UFPA	Universidade Federal do Pará
VQM	Video Quality Metric
WMNs	Wireless Mesh Networks

Lista de Figuras

Figura 1	Esquema genérico do uso da codificação de vídeo em redes de computadores.	7
Figura 2	Diferentes níveis de movimentação de um vídeo.	9
Figura 3	Técnica de compensação de movimentação.	9
Figura 4	Nível de complexidade de um vídeo.	11
Figura 5	Ordenamento linear dos coeficientes DCT de um macrobloco 8x8.	12
Figura 6	Estrutura hierárquica dos <i>frames</i> do MPEG.	13
Figura 7	Diferentes escalas da metodologia SS.	17
Figura 8	Visão geral do MultiQoE.	34
Figura 9	Funcionamento do MultiQoE.	35
Figura 10	Etapas de implementação do MultiQoE.	37
Figura 11	Vídeos selecionados.	38

Figura 12	Árvore das distâncias Euclidianas.	39
Figura 13	Rede em malha sem fio da UFPA.	41
Figura 14	Taxa de perdas de pacotes em todos os experimentos.	41
Figura 15	Histograma das notas obtidas.	43
Figura 16	A influência do GoP na composição do fluxo de vídeo.	44
Figura 17	Composição do <i>Backbone</i> da rede em malha sem fio da UFPA.	48
Figura 18	MultiQoE vs PSQA.	49
Figura 19	MultiQoE vs PSNR.	50
Figura 20	MultiQoE vs SSIM.	50
Figura 21	MultiQoE vs VQM.	51
Figura 22	MSE para cada cluster.	51
Figura 23	MSE para cada tamanho de GoP.	52

Lista de Tabelas

Tabela 1	Mapeamento do PSNR para o MOS	20
Tabela 2	Trabalhos relacionados.	32
Tabela 3	Tamanho médio em bits dos coeficientes DCT e dos vetores de movimentação dos vídeos	38
Tabela 4	Características e classificação dos vídeos selecionados	40
Tabela 5	Influência do conteúdo na composição do fluxo de vídeo.	44
Tabela 6	Parâmetros de simulação configurados no simulador.	48

CAPÍTULO 1

Introdução

1.1 Qualidade de experiência em redes em malha sem fio

Segundo o *Cisco's Visual Networking Index* [CVNI 2010], espera-se que em 2020, os serviços baseados em transmissão de vídeo representarão cerca de 90% de todo tráfego da Internet.

Um dos motivos para o crescimento extraordinário das aplicações de vídeo são os usuários finais, os quais estão sempre em busca de novas aplicações que permitem maior interatividade e percepção visual, deixando de lado as aplicações “estáticas”.

Os serviços de texto plano tal como mensagem instantânea vêm perdendo lugar para aplicações que envolvem áudio e vídeo, por exemplo, os jogos interativos que antes eram executados localmente na máquina do usuário, hoje são executados de forma interativa entre os usuários de uma rede (local ou Internet).

Pode-se citar também o sucesso de sites que disponibilizam o serviço de conteúdo de vídeo armazenado e a notável popularização do serviço de armazenamento de vídeo na Internet, onde os usuários criam, compartilham e acessam milhares de vídeos diariamente.

Outra contribuição para o crescimento das aplicações multimídia (áudio e vídeo) são os grupos corporativos. Empresas, universidades, associações e órgãos públicos vem investindo na implantação e disponibilização de serviços de teleconferência, telemedicina, ensino à distância, VoIP (*Voice over Internet Protocol*) e muitos outros.

Além disso, a busca dos usuários finais e grupos corporativos por novas tecnologias propiciaram a popularização das redes sem fio que oferece redução de custo e mobilidade aos usuários. Assim, as redes sem fio estão cada vez mais presentes como

tecnologia de acesso de primeira milha, fornecendo acesso à Internet em praças, hotéis, aeroportos, empresas e residências. Esta tendência vem crescendo ainda mais devido à redução de custos de equipamentos portáteis como *notebooks*, *netbooks*, celulares e outros dispositivos que possuem interfaces sem fio.

Porém, apesar dos muitos atrativos, as redes sem fio estão sujeitas a alto nível de ruído, interferência, queda de enlaces, enfraquecimento do sinal e outros problemas. Diversas soluções são constantemente pesquisadas e testadas a fim de contornar tais problemas, fazendo existir neste tipo de tecnologia uma dinâmica muito grande de inovações.

Atualmente existem muitas tecnologias sem fio, e dentre tais tecnologias, as redes em malha sem fio (*Wireless Mesh Networks* - WMNs) vem se destacando por apresentar infraestrutura de baixo custo, alta confiabilidade e escalabilidade. Além disso, as WMNs possuem as características de serem auto-configuráveis, auto-organizáveis e de rápida implantação [Akyildiz and Wang 2009].

Antes da era de comunicação multimídia, para avaliar a qualidade de serviços tradicionais como *e-mail*, navegação *web* e transferência de arquivos, era suficiente analisar parâmetros simples de Qualidade de Serviço (*Quality of Service* - QoS), uma vez que a definição de qualidade destas aplicações não possuía complexidade [Piamrat et al. 2009b]. Estes parâmetros se resumem basicamente em largura de banda, perda de pacotes, atraso e variação de atraso.

Porém, para avaliar a qualidade das aplicações multimídias, os parâmetros de QoS não são suficientes para refletir a percepção humana. Desta forma, fez-se necessário o surgimento de novos parâmetros de qualidade baseados nas propriedades e atitudes relevantes que os seres humanos possuem ao avaliar um serviço (vídeo, áudio, imagem etc). Assim o conceito de Qualidade de Experiência (*Quality of Experience* - QoE) surgiu como sendo o nível de aceitação geral da aplicação ou como o usuário final percebe subjetivamente o serviço [ITU-TP.10/G/100 2007].

No entanto, estimar o nível de QoE de uma aplicação multimídia em tempo real é uma tarefa complexa visto que deve levar em consideração diversos parâmetros que vão desde as características do sistema visual humano até erros de codificação que distorcem a imagem. Este desafio se torna ainda maior quando a aplicação de vídeo utiliza um ambiente em malha sem fio de múltiplos saltos, pois há maior probabilidade que os pacotes sofram excesso de atraso e a qualidade de transmissão seja afetada devido ao tempo excessivo gasto para que os pacotes cheguem até o destinatário.

1.2 **Objetivos**

O objetivo geral deste trabalho é propor um mecanismo, denominado MultiQoE, capaz de estimar a qualidade das aplicações de vídeo em tempo real com alta correlação com a qualidade percebida pelo usuário. Além disso o mecanismo proposto nesta dissertação deve satisfazer os seguintes objetivos específicos:

1. Estimar QoE em tempo real levando em consideração não só parâmetros de QoS, mas também parâmetros de codificação.
2. Utilizar parâmetros de *codec* que reflitam os diversos impactos que podem ocorrer aos vídeos com diferentes tipos de conteúdo.
3. Estimar QoE para vídeos com diferentes codificação MPEG.
4. Realizar a classificação do conteúdo de vídeo quanto ao nível de atividade espaço-temporal.
5. Considerar condições de rede mais próximas às apresentadas em cenários reais.

1.3 Organização do texto

O restante do documento está dividido em 6 capítulos. O Capítulo 2 apresenta conceitos relevantes sobre as tecnologias utilizadas na compressão de vídeos para sistemas multimídia, ressaltando um breve histórico da padronização MPEG, além de detalhar a estrutura do MPEG. O Capítulo 3 apresenta a classificação dos métodos e métricas que estimam a qualidade de aplicações de vídeo. Também são apresentados no Capítulo 3 os principais conceitos sobre QoE e as diferenças entre QoE e QoS.

No Capítulo 4 há a descrição dos trabalhos que propõem mecanismos capazes de fornecer uma estimativa de QoE em tempo real. O Capítulo 5 apresentará a visão geral, o funcionamento e as etapas de implementação do método de estimativa chamado MultiQoE. O Capítulo 6 apresenta um estudo de caso do MultiQoE em uma rede em malha sem fio, detalhando as configurações do ambiente e os resultados obtidos. Por fim, o Capítulo 7 descreve as considerações finais sobre o trabalho e também os trabalhos futuros.

CAPÍTULO 2

Compressão de vídeo para sistemas multimídia

Este capítulo tem o objetivo de fornecer conceitos sobre tecnologias de compressão de vídeo, os quais posteriormente serão importantes para o entendimento de como vídeos podem sofrer diferentes níveis de degradação ocasionados pela perda de pacotes em uma rede de computadores. A Seção 2.1 introduz conceitos importantes sobre tecnologias de compressão de vídeo e um breve histórico da padronização MPEG. A Seção 2.2 apresenta as técnicas de compressão temporal e espacial. A Seção 2.3 apresenta os tipos de *frames* do MPEG.

2.1 Introdução à compressão de vídeo

Entre as aplicações da Internet, os serviços que envolvem a transmissão de vídeo em tempo real tem se tornado cada vez mais popular, como por exemplo transmissão de eventos esportivos, noticiários, vídeoconferência, telemedicina, aplicações militares e vigilância remota. Tais aplicações têm atraído milhões de usuários ao redor do mundo e gerado um significativo montante de dados sobre a Internet.

As aplicações de vídeo funcionam exibindo sucessivamente um conjunto de *frames* (quadros) ao usuário. Os *frames* são exibidos por um dado período e assim, no cérebro humano, cria-se uma ilusão de movimento. É importante destacar que definiu-se neste trabalho que a palavra *frame* é um sinônimo de quadro.

Atualmente existem vários cenários de redes de computadores para transmissão de vídeo em tempo real, porém o tipo de sistema mais utilizado ainda é formado por servidores de vídeos que disponibilizam conteúdo aos clientes, os quais representam os

consumidores deste serviço, requisitando vídeos de sua preferência. Neste contexto, a principal função das redes de computadores é realizar o transporte de dados entre os clientes e servidores [Zhang and Hassanein 2010].

No entanto, a Internet não foi projetada, originalmente, para suportar transmissões de dados em tempo real, pois na época de sua criação, as aplicações predominantes (páginas da *web*, *e-mail*, transmissão de arquivos de texto etc) não possuíam requisitos restritos de qualidade.

Com o surgimento das aplicações de vídeo, os algoritmos de compressão ganharam destaque visto que estes removem redundâncias temporais e espaciais do vídeo. Além da redução da quantidade de dados necessários para representar um vídeo (espaço de armazenamento), a compressão de vídeo possibilitou um melhor aproveitamento da banda da rede, a qual é um fator limitante para a maioria dos usuários da Internet [Hanzo et al. 2007].

As famílias mais importantes de algoritmos de compressão de vídeo estão publicadas e reconhecidas pelos seguintes órgãos: *International Organization for Standardization* (ISO) e *International Telecommunication Union* (ITU).

O *Motion Picture Expert Group* (MPEG) surgiu em 1988 com a junção do ISO e o comitê técnico do *International Electrotechnical Commission* (IEC). O objetivo de sua criação foi desenvolver uma padronização para codificação de áudio e vídeo. O primeiro padrão foi chamado de MPEG-1 e lançado em 1992. Em 1994 o MPEG-2 foi lançado com o objetivo de ampliar o escopo de atuação do MPEG-1. Nesta nova versão foram acrescentadas especificações para a codificação de áudio e vídeo para diferentes tipos de aplicações áudio-visuais.

O MPEG-4 foi iniciado em 1994 com o objetivo de criar padrões de codificação de áudio e vídeo para ambientes que estejam propensos a erro, exijam alto grau de compressão e interatividade. Encaixam-se neste tipo de ambiente as redes de computadores.

Os padrões de codificação desenvolvidos pelo MPEG foram mundialmente adotados pela indústria de equipamentos de áudio e vídeo, tornando-se um fator chave para o sucesso da padronização MPEG. Além disso, a grande abrangência do padrão e a possibilidade de suportar uma vasta gama de aplicações áudio-visuais foram essenciais para a grande aceitação do MPEG pelo mercado e usuários.

Como apresentado na Fig. 1, o MPEG-4 atua em dois momentos durante a transmissão de um vídeo. No remetente, o codificador (*Codec*) tem a função de remover as informações redundantes e codificar o vídeo em um formato intermediário que será transmitido pela rede. Do lado do receptor, o decodificador (*Decodec*), transformando o formato intermediário em vídeo [Gualdi et al. 2008].

O arquivo intermediário criado pelo codificador é uma conjunto de bits e a sintaxe destes dados é definida pelo MPEG-4. É através desta sintaxe que o decodificador consegue reconstruir o vídeo original. A padronização do MPEG-4 fornece poucas informações específicas sobre o modo de operação do codificador, visto que a intenção é deixar que

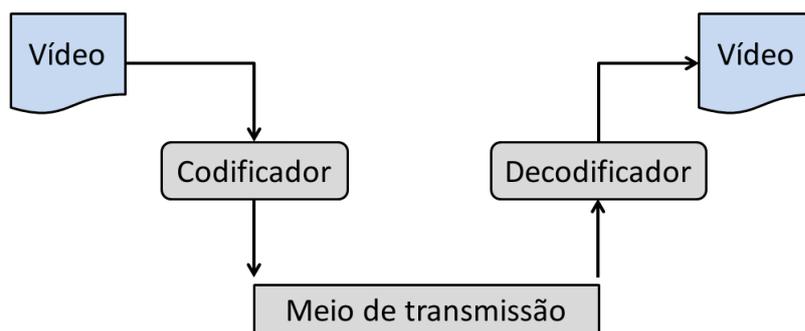


Figura 1: Esquema genérico do uso da codificação de vídeo em redes de computadores.

cada fabricante possua sua própria versão de codificador.

A especificação do MPEG concentra-se principalmente no formato da corrente de bits que os codificadores devem produzir. Agindo dessa forma, o padrão MPEG garante que qualquer decodificador MPEG seja capaz de decodificar vídeos codificados por codificadores de diferentes fabricantes, independentemente do modo como o codificador esteja implementado. A vantagem de deixar com que cada fabricante possa desenvolver sua própria versão de codificador é permitir que cada instituição possa utilizar seus próprios algoritmos. A única exigência é que a corrente de bits gerada mantenha compatibilidade com o padrão determinado pelo MPEG.

2.2 Princípios de compressão de vídeo

No contexto de sistemas computacionais, para comprimir uma dada informação é necessário reduzir o número de dados exigidos para representá-la. Como as informações são representadas por conjuntos de bits, a tarefa da compressão é representar uma dada informação com a menor quantidade de bits [Hanzo et al. 2007].

Um princípio importante para redução de bits é analisar o grau de preeditibilidade da informação em questão. Por exemplo, a função seno pode ser preedita com alto grau de acerto porque esta função se repete em todos os ciclos. Por outro lado, um sinal de ruído possui baixo grau de preeditibilidade, por não haver ciclos periódicos neste tipo de sinal. No caso específico de vídeo, a taxa de informação presente nos vídeos é algo variável e é um elemento difícil de ser preedito.

Neste contexto, a definição de entropia de uma amostra de vídeo é dada em função de quanto o conteúdo de tal amostra é diferente do valor esperado [Watkinson 2004]. Dessa forma, quanto mais entropia uma amostra possui, menor é o grau de compressão, visto que o codificador pode prever menos informações da amostra. Por outro lado, quanto menor é a entropia de um vídeo, maior é o grau de compressão da amostra.

Como abordado anteriormente, durante a execução de um vídeo, a taxa de informação presente no conteúdo de tal vídeo é variável. De forma similar, a entropia do vídeo também é variável, fazendo com que a taxa de compressão varie ao longo da

execução do vídeo.

Assim, o cenário ideal para a transmissão de vídeos comprimidos seria através de canais que fossem capazes de fornecer a largura de banda necessária segundo a taxa de compressão do vídeo. Isto introduz um fator crítico para aplicações de transmissão de vídeo em tempo real sobre redes de computadores, uma vez que segundo o modelo de melhor esforço, no qual se baseia a Internet, não há garantias de banda e isso em alguns casos pode ser insuficiente para a transmissão de vídeo comprimido.

As Seções 2.2.1 e 2.2.2 abordarão com maiores detalhes como o MPEG-4 reduz as redundâncias espaciais e temporais, respectivamente.

2.2.1 Compressão temporal

A compressão temporal é baseada na técnica de compensação de movimentação. Esta técnica reduz as redundâncias existentes em *frames* consecutivos e o principal objetivo é criar um *frame* a partir dos *frames* anteriores, eliminando as partes em comum dos *frames*. Assim, quando um vídeo está sendo transmitido, não é necessário enviar cada *frame* inteiro, mas apenas as diferenças entre os *frames*. As Figs. 2.a e 2.b exemplificam essa ideia apresentando dois *frames* consecutivos de um vídeo.

Se a seguinte subtração for feita: *frame 2* - *frame 1*, é possível observar o *frame*-diferença obtido na Fig. 2.c, onde a porção em preto representa as partes em comum aos dois *frames*. Porém, as demais partes representam a variação entre os *frames*. Dessa forma, é possível reconstruir o *frame 2* a partir do *frame 1* e do *frame*-diferença. De forma similar, a Fig. 2.f apresenta o *frame* obtido a partir das Fig 2.d e 2.e. É possível observar que a Fig. 2.f possui bem menos partes em preto do que a Fig. 2.c, indicando que este vídeo apresenta um nível maior de movimentação. Além disso, analisando as Fig. 2.f e 2.c é possível concluir que a taxa de compressão temporal é menor para vídeos com alta movimentação.

Apesar da ideia básica da compensação de movimentação do MPEG ser a de construir um *frame* a partir do anterior, o MPEG aplica o conceito de compensação de movimentação individualmente a cada macrobloco, o qual consiste em partes menores do *frame*. Isto significa que um macrobloco é criado a partir do seu macrobloco referência, não sendo necessário transmitir as partes em comum dos dois macroblocos.

Supondo a existência dos *frames* X, Y e de um macrobloco A que pertence ao *frame* X, onde o *frame* X é exibido no instante T-1 e o *frame* Y é exibido no instante T. Para diminuir a quantidade de dados necessários para representar o vídeo, o macrobloco do *frame* Y que possuir maior similaridade com o macrobloco A deverá ser escolhido para ser referência do macrobloco A. No momento da decodificação, cada macrobloco construído sob o conceito de compensação de movimentação deve possuir a localização de seu macrobloco referência. A localização do macrobloco referência é armazenado em um campo do MPEG denominado de vetor de movimentação.



Figura 2: Diferentes níveis de movimentação de um vídeo.

A Fig. 3 apresenta a utilização do vetor de movimentação na compressão temporal. É possível observar que no *frame* N há um objeto localizado na parte inferior da tela. No entanto no *frame* $N+1$ o objeto localiza-se na parte superior da tela. Após realizar uma busca no *frame* N é possível obter o vetor de movimentação representado na Fig. 3 pela seta em vermelho. A origem da seta indica o macrobloco referência e a ponta da seta indica o macrobloco a ser criado a partir de sua referência.

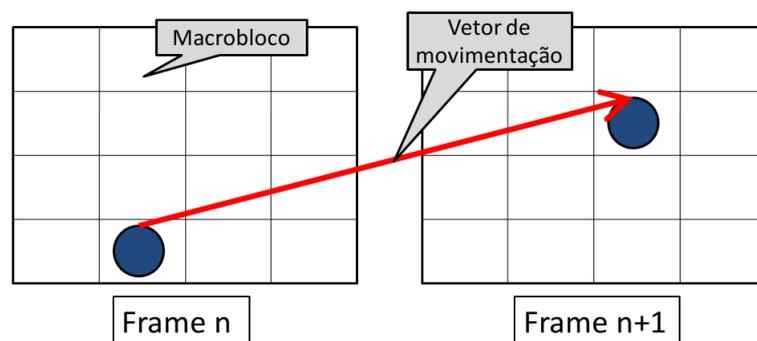


Figura 3: Técnica de compensação de movimentação.

A utilização da técnica de compressão temporal, onde envia-se os macroblocos diferença, há vantagens e desvantagens. A grande vantagem é a transmissão apenas das novas informações presentes nos macroblocos futuros, havendo a redução de banda para transmissão de um vídeo. A desvantagem é dependência entre os *frames*.

Por outro lado, caso haja uma perda que afete macroblocos referência, o erro será propagado pelos próximos macroblocos. Para amenizar o problema da propagação de erros devido ao uso de macroblocos referência, o MPEG-4 é projetado para enviar periodicamente *frames* constituídos de macroblocos que não necessitam de referência para serem decodificados. Os tipos de *frames* do MPEG serão abordados com maiores detalhes na seção 2.3.

2.2.2 Compressão espacial

A compressão espacial é definida como sendo técnicas que removem redundâncias de informações presente nos *frames* baseado na propriedade de que os *pixels* de um *frame* estão relacionados com seus vizinhos [Greengrass et al. 2009].

O primeiro passo para compressão espacial de um *frame* é realizar a análise das frequências de tal imagem para encontrar as frequências dominantes. Análise das frequências é feita passando a imagem para o domínio da frequência através de técnicas denominadas de transformada. A transformada mais conhecida é a transformada de *Fourier* e é capaz de expressar uma imagem no domínio da frequência, fornecendo um conjunto de coeficientes, os quais descrevem o quanto uma dada frequência está presente na imagem.

O processo inverso é feito através da transformada inversa, a qual é capaz de reproduzir a imagem original a partir dos coeficientes. Porém, apesar da transformada de *Fourier* ser largamente conhecida, o MPEG-4 utiliza a transformada DCT (*Discrete Cosine Transform*) por produzir coeficientes apenas da função cosseno e por isso tornar-se mais prática [Watkinson 2004].

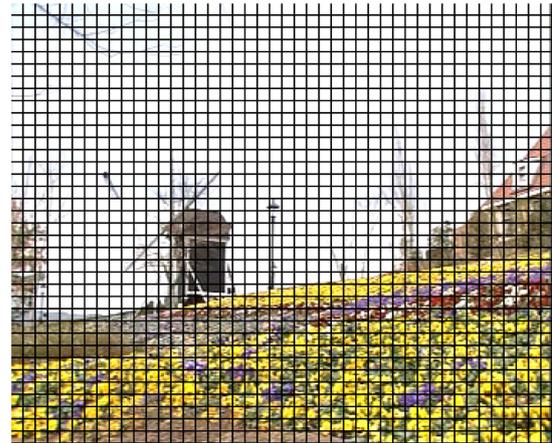
As Figs. 4.a e 4.b apresentam um *frame* dos vídeos *Hall* e *Flower*, respectivamente. Além disso, é possível observar em cada Figura a marcação dos macroblocos. A divisão de um *frame* em macroblocos ocorre pois (como explicado anteriormente) a técnica de compressão é feita para cada macrobloco em particular.

O tamanho de um macrobloco é variável, porém o tamanho padrão é uma matriz de 8x8 *pixels*. Especificamente na compressão espacial, a transformada DCT é aplicada a cada macrobloco. A transformada produz um coeficiente para cada *pixel* do macrobloco, assim se um macrobloco possui 64 *pixels* (8x8), a transformada produzirá uma matriz 8x8 contendo 64 coeficientes.

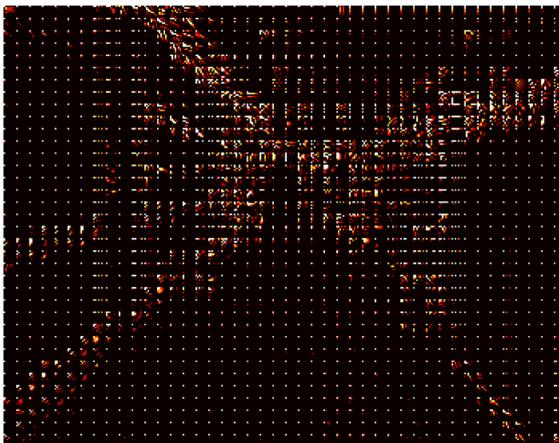
Para exemplificar esse processo as Figs. 4.c e 4.d apresentam os macroblocos obtidos após a aplicação da transformada nas Figs. 4.a e 4.b, respectivamente. Para ficar com maior percepção visual, o valor dos coeficientes está apresentado segundo uma escala de cores onde o preto representa valores de coeficientes mais baixos, enquanto a cor branca representa os coeficientes de valores maiores. Como há um domínio de macroblocos mais próximos de branco na Fig. 4.b, esta Figura deixa clara a existência de coeficientes de valores maiores na Fig. 4.b do que na Fig. 4.e.



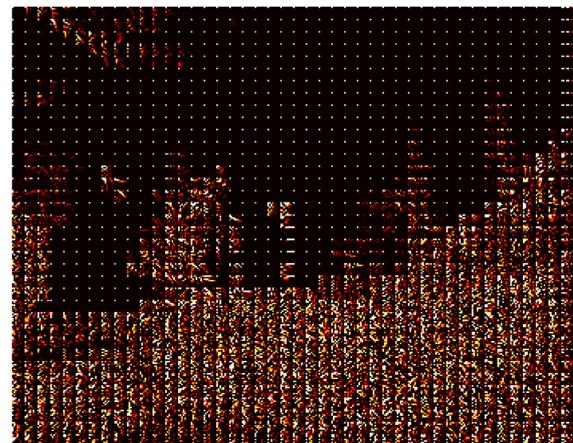
a) Divisão dos macroblocos de um Frame do vídeo Hall.



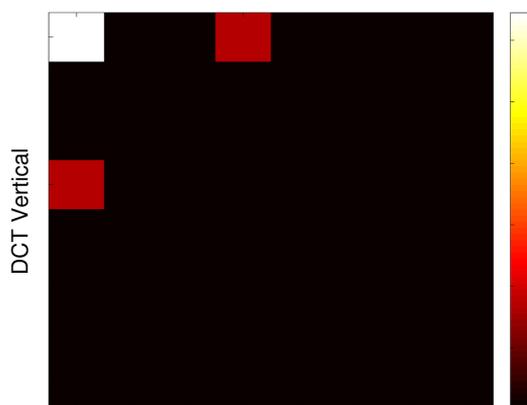
b) Divisão dos macroblocos de um Frame do vídeo Flower.



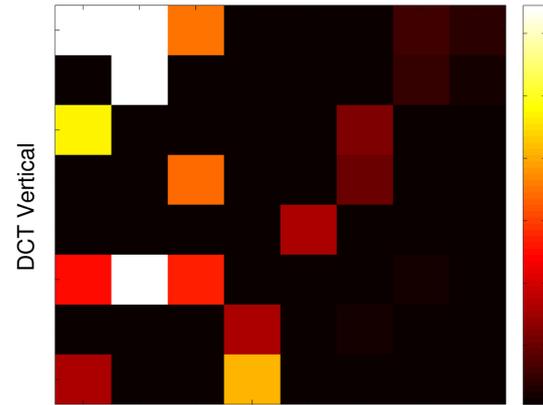
c) Representação dos coeficientes dos macroblocos do Frame do vídeo Hall.



d) Representação dos coeficientes dos macroblocos do Frame do vídeo Flower.



e) Único macrobloco do vídeo Hall.



f) Único macrobloco do vídeo Flower.

Figura 4: Nível de complexidade de um vídeo.

As Figs. 4.e e 4.f apresentam os coeficientes horizontais e verticais de dois macroblocos em particular das Figs. 4.c e 4.d, respectivamente. Na matriz contendo os coeficientes de um macrobloco, a frequência cresce da esquerda para direita e de cima

para baixo como apresenta a Fig. 5. Através dessa ordem o MPEG dispõe os coeficientes em um vetor na ordem crescente de frequência.

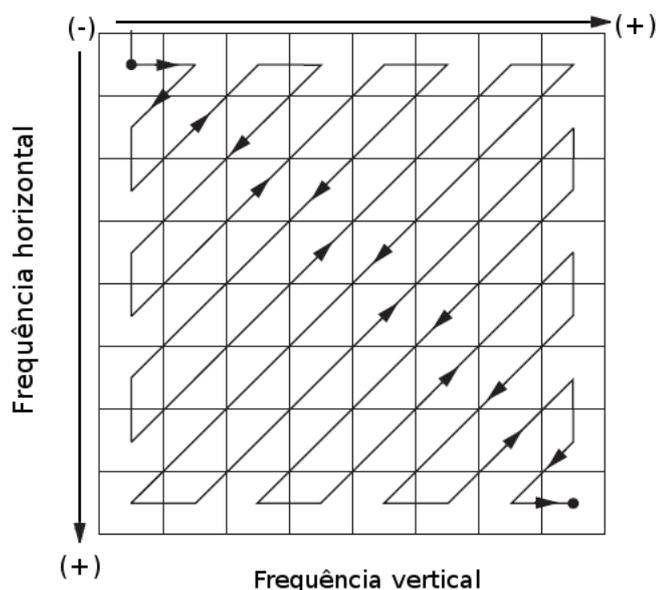


Figura 5: Ordenamento linear dos coeficientes DCT de um macrobloco 8x8.

A compressão espacial é dada pela redução do número de coeficientes transmitidos pois aqueles coeficientes das frequências mais altas que possuem valores próximo de zero podem não ser transmitidos sem afetar muito a qualidade de vídeo, havendo assim uma redução do número de bits utilizado para representar um *frame*. Porém o nível de compressão espacial dado pela transformada DCT é variável visto que depende das frequências da imagem, por exemplo, a Fig. 4.b apresenta mais coeficientes de valor baixo para frequências altas do que os coeficientes da Fig. 4.e. Isso nos permite concluir que baseado no número de coeficientes DCT presente em um vídeo MPEG é possível inferir o nível de compressão espacial do vídeo.

2.3 Estrutura do MPEG-4

O MPEG define sua estrutura baseada em três tipos de *frames*, denominados: I (Intra), P (Preditivo) e B (Bi-direcional). A compressão espacial é aplicada aos macroblocos dos *frames* do tipo: I, P e B, porém a compressão temporal é aplicada somente aos macroblocos dos *frames* do tipo: P e B. Como apresenta a Fig. 6, os macroblocos dos *frames* do tipo P utilizam como referência macroblocos dos *frames* anteriores do tipo I ou P, já os macroblocos dos *frames* do tipo B utilizam como referência macroblocos dos *frame* anteriores ou futuros do tipo I ou P.

Um fator muito importante da codificação do MPEG é o tamanho do GoP (*Group of Picture*). Tal comprimento indica a frequência com que os *frames* I existirão no vídeo codificado, por exemplo, um vídeo codificado com GoP de comprimento 10 implica que os *frames* I irão estar presentes a cada 10 *frames* do vídeo. Além do tamanho do GoP

existem diferentes padrões internos, os quais definem o número de *frames* P para cada *frame* B. Para exemplificar a estrutura de um GoP de vídeo MPEG, a Fig. 6 apresenta uma codificação de vídeo onde o tamanho e o padrão interno do GoP são, respectivamente, 10 e de um *frame* P para dois *frames* B.

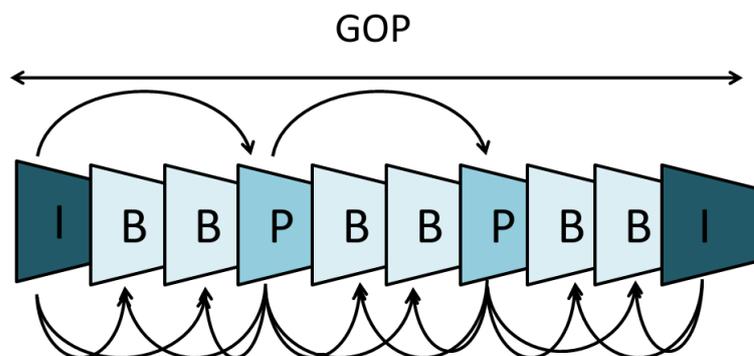


Figura 6: Estrutura hierárquica dos *frames* do MPEG.

A principal consequência do fato do *frame* I não realizar compressão temporal é que quando um erro afetar *frames* do tipo B ou P, esses erros serão propagados pelos demais *frames* até que um *frame* I chegue até o receptor, ou seja, o erro será propagado por todo GoP. No caso da perda em um *frame* P, o erro será propagado pelos *frames* P e B restantes no GoP. E caso um erro ocorra em um *frame* B, este erro não será propagado visto que os *frames* B não são usados como referência por outros *frames*.

2.4 Conclusão do Capítulo

Este capítulo apresentou tecnologias, conceitos e aplicações relevantes às aplicações de vídeos que realizam a transmissão de vídeo em tempo real em meios de comunicação sujeitos a erros/perdas. O padrão MPEG de codificação de vídeo mereceu maior destaque visto que se tornou um padrão de codificação de vídeos muito utilizado pela indústria e pelos usuários que produzem, compartilham e consomem serviços de vídeos na Internet.

O padrão MPEG utiliza técnicas de compressão temporal e espacial durante o processo de codificação do vídeo. Além disso, o MPEG define três tipos de *frames*: I, P e B. Os *frames* I fazem uso apenas da compressão espacial. Já os *frames* P e B utilizam compressão espacial e temporal.

CAPÍTULO 3

Qualidade de Experiência

Qualidade de Experiência (*Quality of Experience* - QoE) tem se tornado crucial para avaliação da qualidade de aplicações de vídeo e por isso este capítulo tem por objetivo apresentar as principais definições e abordagens de QoE. A Seção 3.1 apresenta os principais conceitos sobre QoE assim como as diferenças existentes entre QoE e Qualidade de Serviço (*Quality of Service* - QoS). A Seção 3.2 descreve as principais metodologias subjetivas e as métricas objetivas e híbridas de estimativa de QoE.

3.1 Introdução à Qualidade de Experiência

Qualidade de Experiência de um serviço de vídeo é definida como a aceitabilidade geral da aplicação de vídeo ou como o usuário final percebe subjetivamente o serviço [ITU-TP.10/G/100 2007]. Este conceito surgiu como uma extensão aos antigos parâmetros de QoS, visto que QoS mensura a qualidade das aplicações apenas através do desempenho dos equipamentos da rede e isso não revela claramente com qual grau de qualidade o usuário percebe a aplicação.

Atualmente, o nível de qualidade das aplicações de vídeo que são transmitidas em redes de computadores depende largamente da capacidade do desempenho do codificador de vídeo, da largura de banda, do atraso, *jitter* e taxa de perda. Porém todos esses fatores de degradação podem ser agrupados, basicamente, em duas fases, são elas: codificação e transmissão.

Nesta direção, várias pesquisas estão sendo feitas para mensurar o nível de qualidade percebido das aplicações de vídeo, focando-se em quantificar a degradação que é inserida pelo processo de codificação e/ou a transmissão do vídeo codificado em canais com limitação de banda [Koumaras et al. 2010].

Quanto à primeira fase, sabe-se que a degradação de qualidade devido ao processo de codificação está fortemente relacionada aos parâmetros selecionados para codificação do vídeo. Tais parâmetros podem ser alterados de acordo com o nível de compressão que se deseja atingir, o qual será determinante na ocorrência de perda no nível de qualidade das imagens exibidas pelo vídeo.

Concernente ao processo de transmissão de vídeo, é necessário considerar que os serviços de compressão de vídeo tornam as aplicações de vídeo extremamente sensíveis à perdas de dados, devido à estrutura de interdependência existente entre os *frames* de um vídeo comprimido. Dessa forma, as aplicações de vídeo passam a exigir alto grau de confiabilidade dos meios de transmissão. Porém, o atual modelo de melhor esforço, no qual se baseia a Internet, não fornece garantias dos requisitos mínimos das aplicações de vídeo, tornando imprevisível a degradação ocasionada pela transmissão na Internet.

Assim, a tarefa de mensurar a degradação da qualidade percebida pelo usuário é extremamente complexa por lidar com alto grau de subjetividade, sistema visual humano, tecnologias de codificação de vídeo e fatores de limitação da rede de transmissão.

3.2 Métodos e métricas de estimativa de QoE

Há diversas formas de estimar a qualidade de um vídeo. É possível classificar as propostas que estimam a qualidade de vídeo em: subjetiva, objetiva e híbrida. Porém, faz-se necessário estabelecer a diferença entre método e métrica. Neste trabalho, optou-se por considerar *método* as soluções subjetivas e híbridas, uma vez que exigem procedimentos rigorosos. Por outro lado, a estimativa de QoE de forma objetiva é dada pela execução de algoritmos e/ou fórmulas matemáticas, por isso estas soluções recebem a denominação de métricas.

As métricas objetivas podem ser caracterizadas em: referência completa (*Full Reference* - FR), referência reduzida (*Reduced Reference* - RR) e sem referência (*No Reference* - NR).

As métricas objetivas FR são técnicas de estimativa de QoE que necessitam do vídeo original (não degradado) como referência durante o processo de estimativa de qualidade e devido a esta característica as métricas FR não podem ser utilizadas para estimar QoE de aplicações de vídeo transmitidos em tempo real em redes de computadores, visto que as métricas FR necessitam do vídeo degradado (transmitido) e não degradado (original) simultaneamente.

Devido à esta razão, pesquisas desenvolveram métricas que extraem características estruturais do vídeo original (RR) ou não necessitam de nenhuma informação do vídeo não degradado (NR), como por exemplo [Yamada et al. 2007] e [Gunawan and Ghanbari 2008]. Porém, apesar do desenvolvimento das métricas RF e NF, as métricas FR ainda são largamente utilizadas pela comunidade científica, visto que as métricas RR e NR sofrem pela falta de detalhamento e disponibilização de seus algoritmos. Devido a isto, este trabalho

restringiu-se a utilizar e descrever apenas as principais métricas objetivas FR.

Tendo em vista a classificação dos métodos e métricas de estimativa de QoE, as Seções 3.2.1 e 3.2.3 apresentarão os principais métodos subjetivos e híbridos, respectivamente. Além disso, a Seção 3.2.2 apresentará as principais métricas objetivas.

3.2.1 Métodos Subjetivos

Os métodos de avaliação de qualidade subjetiva são aqueles baseados nas notas de qualidade dado por um avaliador humano após ter assistido um vídeo. A qualidade final de um vídeo avaliado subjetivamente é dada por uma média de avaliações subjetivas denominada de *Mean Opinion Score* (MOS). Portanto, existe uma relação entre os métodos subjetivos e o MOS, onde o primeiro é o processo ou a forma que os vídeos serão avaliados subjetivamente, e o segundo é o objetivo final de todos os métodos subjetivos.

O MOS é o modo que reflete de maneira mais clara a percepção do usuário, visto que a estimativa de qualidade do vídeo é obtida diretamente pela avaliação humana, levando em consideração os fatores subjetivos de qualidade existentes em cada avaliação. No entanto, a principal desvantagem do MOS é que as metodologias de avaliação subjetiva impossibilitam seu funcionamento em tempo real, visto que esta exige muitos recursos (humanos, tempo) durante o processo de realização dos testes subjetivos.

Na comunidade científica há especificações bem definidas para realização das avaliações subjetiva de qualidade: [ITUR-B500-11 2002] e [P.900 2002]. Tais especificações são de extrema importância para que diferentes grupos de pesquisa possam trocar informações sobre os resultados dos testes subjetivos de qualidade e para que haja um controle de qualidade nos testes subjetivos aplicados por diversos pesquisadores.

Existe uma simples classificação dos métodos de avaliação subjetiva: avaliação qualitativa e avaliação quantitativa. As avaliações qualitativas existem para descrever a percepção de qualidade dos humanos. Tal método de avaliação subjetiva é centrada em descrever os fatores humanos que influenciam os avaliadores durante os testes subjetivos de qualidade. Esta metodologia é utilizada em trabalhos que explicam os aspectos sociológicos e psicológicos da qualidade [Bouch and Sasse 2001].

A metodologia de avaliação quantitativa é mais utilizada em contextos das ciências exatas, visto que este tipo de avaliação de qualidade gera valores que representam a qualidade de um vídeo/imagem e por isso esta metodologia é mais útil para a proposta do presente trabalho, sendo que maiores detalhes das metodologias quantitativa serão apresentado a seguir.

3.2.1.1 Single Stimulus (SS)

Esta metodologia de avaliação é denominada de *Single Stimulus* (SS) na recomendação [P.900 2002] e de *Absolute Category Rating* (ACR) em [ITUR-B500-11 2002].

De forma geral nesta metodologia, um vídeo é apresentado a um avaliador humano e após a apresentação do vídeo, o avaliador é questionado para fornecer uma nota que reflita a qualidade geral do vídeo apresentado. Um único vídeo é apresentado por vez e o tempo máximo para o avaliador fornecer a nota é estipulado em 10 segundos. A ordem dos vídeos apresentados necessita ser aleatória e preferencialmente deve haver uma ordem diferente para cada observador.

Existem muitas variações da metodologia SS. Algumas delas são: apresentar o mesmo vídeo uma única vez ou múltiplas vezes para o mesmo avaliador em momentos diferentes da avaliação, utilizar duas ou três quantidades de telas em cinza em diferentes momentos da apresentação do vídeo e apresentar diferentes tipos de escalas de votação como apresenta a Fig. 7. É possível observar que a Fig. 7.a apresenta a escala com 5 diferentes níveis de qualidade, porém a escala de qualidade presente na Fig. 7.c lista 5 níveis de danos. Por outro lado as Fig. 7.b, Fig. 7.f e Fig. 7.e apresentam escalas contendo 9, 11 e 7 níveis de qualidade, respectivamente. Por fim a Fig. 7.d apresenta uma escala de qualidade contínua onde é possível atribuir uma nota entre 100 e 0.

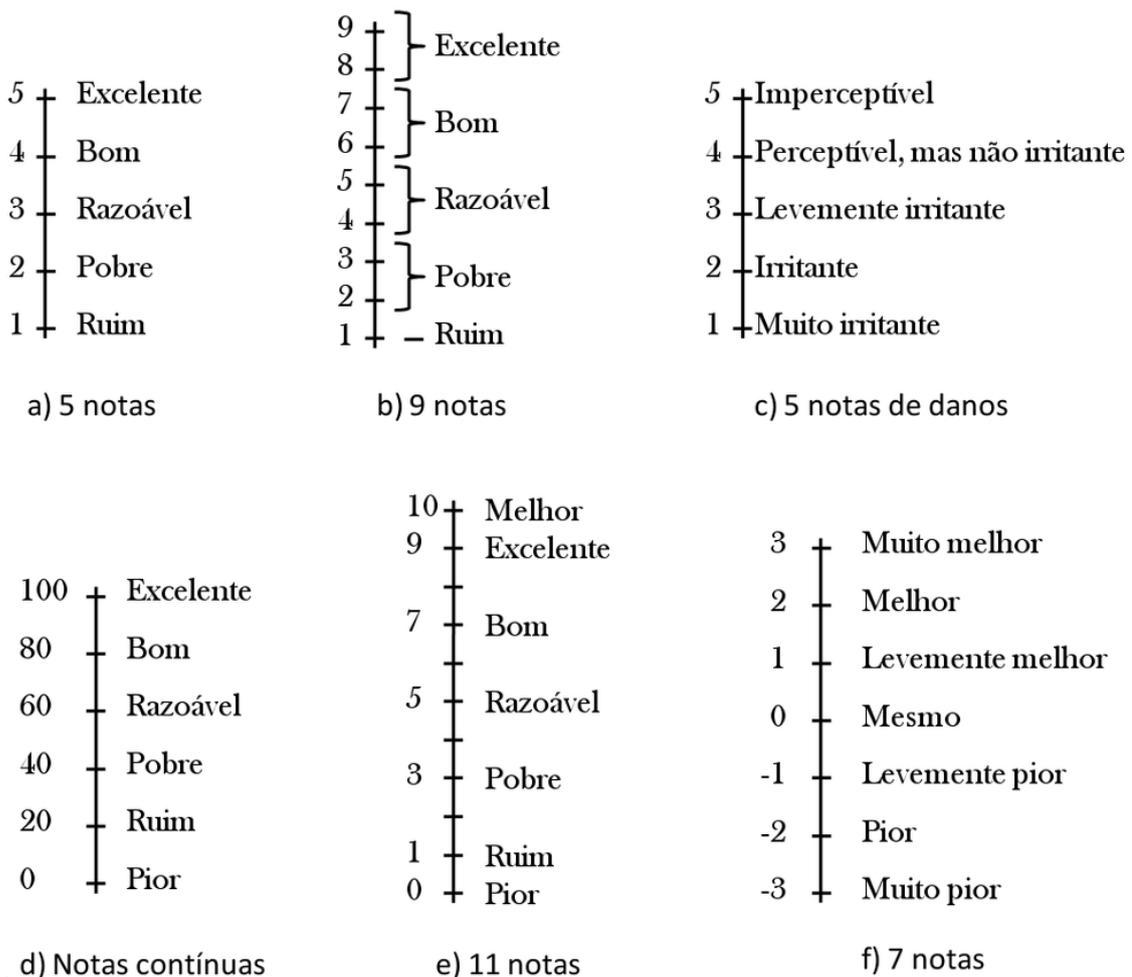


Figura 7: Diferentes escalas da metodologia SS.

Apesar de existir esta grande variedade de escalas de notas de qualidade, todas

as escalas mantêm um certo grau de compatibilidade entre elas. Tais escalas existem para permitir que a avaliação subjetiva possa ser adaptada de acordo com o nível de precisão que se queira obter nas notas subjetivas.

3.2.1.2 Double Stimulus Impairment Scale (DSIS)

Esta metodologia de avaliação é denominada de *Double Stimulus Impairment Scale* (DSIS) na recomendação [P.900 2002] e de *Degradation Category Rating* (DCR) na recomendação [ITUR-B500-11 2002]. Nesta metodologia os vídeos são apresentados em pares aos avaliadores. Primeiramente o avaliador assiste ao vídeo sem degradação (original) e em seguida (mesma tela) é apresentado o vídeo degradado. Após a visualização dos dois vídeos, é requisitado que o avaliador forneça a nota para o vídeo degradado. Os vídeos degradados são apresentados ao avaliador de forma randômica.

No começo de cada sessão de avaliação é realizada uma explicação para os avaliadores sobre o tipo de avaliação, a escala de avaliação, a sequência de vídeos e o tempo de votação. Os avaliadores devem fornecer notas baseadas na degradação geral da qualidade do vídeo. Além disso, no início das sessões de avaliação é feito um período de testes para os avaliadores se adaptarem e esclarecer dúvidas sobre a avaliação e as notas provenientes da fase de teste não são levadas em consideração no cálculo final das notas.

3.2.2 Métricas Objetivas

Como citado anteriormente, a metodologia subjetiva possui problemas críticos que impossibilitam o seu uso nos sistemas de transmissão multimídia. Entre os principais problemas da metodologia subjetiva está o alto custo de: recursos humanos e tempo. Dessa forma, as métricas objetivas surgiram principalmente para oferecer uma avaliação da qualidade de vídeo automatizada, buscando reproduzir as avaliações subjetivas.

O método objetivo de avaliação de qualidade de vídeo é aquele que se baseia em algoritmos que processam os sinais do vídeo e levam em consideração parâmetros do codificador. Porém o método objetivo, em sua grande maioria, fornece uma estimativa da qualidade de vídeo comparando simultaneamente todo ou parte do vídeo degradado com parte ou todo vídeo não degradado (original) e esta restrição torna os métodos objetivos inadequados para o uso em tempo real nos sistemas de transmissão multimídia. Outro problema presente no método objetivo é sua baixa correlação com o MOS, tornando este tipo de métrica pouco confiável para expressar a qualidade percebida por avaliadores humanos.

A seguir serão descritas as métricas objetivas: *Peak Signal to Noise Ratio* (PSNR) [Par et al. 2006], *Video Quality Metric* (VQM) [Pinson and Wolf 2004], *Structural Similarity Index Metric* (SSIM) [Wang et al. 2004].

3.2.2.1 Peek Signal to Noise Ratio (PSNR)

A métrica objetiva mais simples e comum para estimar a qualidade de vídeo é o PSNR. Sua formulação é baseada em uma métrica denominada de *Mean Squared Error* (MSE) que pode ser definida na equação 3.1, abaixo:

$$MSE = \frac{1}{F \times M \times N} \sum_{k=0}^{F-1} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [o_k(i, j) - d_k(i, j)]^2$$

Onde:

F : *Frames* do vídeo

$M \times N$: Dimensão em *pixels* do *frame*.

$o_k(i, j)$: Posição do *pixel* no *frame* original.

$d_k(i, j)$: Posição do *pixel* no *frame* recebido.

O PSNR estabelece uma relação em escala logarítmica entre a razão do valor máximo de luminosidade (grau de claridade ou escuridão) para cada *pixel* e o MSE deste respectivo *pixel*. Assim, o valor do PSNR é obtido através da equação 3.1:

$$PSNR = \log_{10} \left(\frac{L^2}{MSE} \right) \quad (3.1)$$

Onde:

L^2 : Valor máximo de luminosidade

É possível observar na equação 3.1 que o PSNR considera o grau máximo de luminosidade estimado para cada *pixel*, por exemplo, se 8 bits estiverem sendo utilizados em cada *pixel* para representar os valores de luminosidade é possível obter 255 (2^8-1) possíveis valores de luminosidade.

Os valores fornecidos de estimativa de qualidade de vídeo fornecido pelo PSNR é dado em dB (decibel) e a Tabela 1 é largamente utilizada para mapear os valores em PSNR para a escala subjetiva de qualidade da Fig. 7.a.

Porém o PSNR desconsidera fatores importantes do sistema visual humano e o crominância dos *pixels*, tornando a estimativa de qualidade fornecida por este método pouco confiável para expressar a percepção humana de qualidade de vídeo. Apesar disso, o PSNR é um método objetivo largamente utilizado por ser simples e exigir baixo custo computacional.

Tabela 1: Mapeamento do PSNR para o MOS

PSNR (dB)	MOS
> 37	5 (Excelente)
31 - 37	4 (Bom)
25 - 31	3 (Regular)
20 - 25	2 (Pobre)
< 20	1 (Péssimo)

3.2.2.2 Structural Similarity Index Metric (SSIM)

A métrica SSIM melhora o tradicional método PSNR considerando fatores do sistema visual humano. O SSIM compara três parâmetros do vídeo original e o vídeo degradado, são eles: luminosidade, contraste e estrutura dos objetos apresentados em cada *frame* do vídeo.

Denominando o vídeo original de x e o vídeo degradado de y , primeiramente o SSIM calcula a média de luminosidade dos vídeos x e y através da equação 3.2:

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.2)$$

Após obter os valores da média de luminosidade dos dois vídeos x, y , a função $l(x, y)$ compara os valores de luminosidade como apresenta a equação 3.3:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (3.3)$$

Onde:

C_1 : Constante incluída para evitar instabilidade quando $\mu_x^2 + \mu_y^2$ assume valores próximos de zero.

O SSIM, na segunda etapa, calcula a intensidade do contraste dos vídeos x e y dada pela equação 3.4 abaixo:

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2} \quad (3.4)$$

Após obter os valores de σ_x , a função $c(x, y)$ compara os valores de contraste, como apresenta a equação 3.5:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (3.5)$$

Onde:

$C_2 = (K_2L)^2$ e $K_2 \ll 1$ para satisfazer às propriedades inerentes a modelagem matemática do SSIM.

A terceira etapa do SSIM estima a estrutura dos objetos normalizando os valores do vídeo pelo seu desvio padrão e a função apresentada na equação 3.6 apresenta como é feita a comparação entre o vídeo original e o vídeo degradado.

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (3.6)$$

Onde:

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y), \quad (3.7)$$

A estimativa final fornecida pelo SSIM é dada pela equação 3.8 apresentada abaixo:

$$SSIM(x, y) = [l(x, y)^\alpha] \cdot [c(x, y)^\beta] \cdot [s(x, y)^\gamma] \quad (3.8)$$

Para simplificar a equação 3.8 considera-se $\alpha = \beta = \gamma = 1$. A estimativa final de qualidade fornecida pelo SSIM é dada por um valor real entre 0 a 1, onde 0 representa a pior qualidade e 1 representa a melhor qualidade possível de um vídeo.

3.2.2.3 Video Quality Metric (VQM)

A métrica VQM foi desenvolvida pelo *Institute for Telecommunication Science* (ITS). Primeiramente o VQM extrai informações do vídeo original e compara com as informações extraídas do vídeo degradado. Dentre as informações analisadas pelo VQM estão características: espaciais, temporais e cromaticidade. A estimativa final de qualidade é fornecida por uma combinação linear dos parâmetros medidos e o impacto que tais parâmetros exercem no sistema visual humano.

Alguns estudos apresentam que o VQM possui boa correlação com as notas subjetivas de qualidade por considerar fatores do sistema visual humano, porém seu alto desempenho restringe-se a vídeos com baixa taxa de bits e não pode ser usado para mensurar a qualidade de sistemas multimídias em tempo real visto que necessita do vídeo original e degradado no mesmo local [Basso et al. 1996].

3.2.3 Métodos Híbridos

Os métodos híbridos têm surgido como uma alternativa para fornecer estimativa de qualidade de vídeo com alta correlação com a avaliação subjetiva e sem necessitar que os vídeos original e degradado estejam simultaneamente na mesma localização. Os métodos híbridos tem por objetivo unir as vantagens dos métodos subjetivos e objetivos, fornecendo estimativa de qualidade de vídeo de forma automatizada.

A abordagem híbrida, em sua maioria, baseia-se em técnicas de inteligência computacional para relacionar parâmetros que causam distorção de qualidade e a percepção dos avaliadores humanos. De modo geral, a tarefa do método híbrido é prever a qualidade da aplicação de vídeo percebida pelo usuário através de parâmetros de rede e *codec*.

O capítulo 5 detalhará a proposta do método híbrido MultiQoE e no capítulo 4 serão descritos os trabalhos relacionados a esta proposta.

3.3 Conclusão do Capítulo

Este capítulo apresentou as principais características das métricas e métodos de estimativa de QoE, as quais são classificadas em objetivas, subjetivas e híbridas. Os métodos subjetivos descrevem como deve ser realizada a avaliação humana dos vídeos e quais procedimentos devem ser seguidos para obtenção do MOS. As métricas objetivas são aquelas que utilizam de fórmulas e/ou algoritmos para fornecer uma estimativa de QoE sem a necessidade das avaliações subjetivas. Por fim, os métodos híbridos são aqueles que, em sua maioria, utilizam técnicas de inteligência computacional para criar mecanismos que capturam aspectos subjetivos da avaliação humana, podendo ser utilizados sem depender do humano para fornecer estimativa de QoE.

CAPÍTULO 4

Trabalhos relacionados

Existem muitos trabalhos que estudam a maneira com que as perdas de pacotes influenciam a qualidade de vídeo, e muitos desses trabalhos propõem mecanismos que melhoram em algum aspecto a qualidade do vídeo percebido pelo usuário, porém apenas um restrito número de trabalhos propõe mecanismos que sejam capazes de fornecer uma estimativa de QoE em tempo real, já que esta tarefa é complexa por existirem muitos fatores que, combinados, afetam a qualidade percebida de um vídeo.

Assim, este capítulo tem por finalidade apresentar mecanismos de estimativa de QoE em tempo real, e para isso, apresentará na Seção 4.1 a metodologia do PSQA (*Pseudo Subjective Quality Assessment*), bem como os trabalhos que propõem adaptações do PSQA. As Seções 4.2, 4.3, 4.4 e 4.5 apresentarão outras propostas para estimativa de QoE em tempo real.

4.1 PSQA

PSQA é um método híbrido que estima QoE através do aprendizado de redes neurais. Basicamente, testes subjetivos são feitos para uma base de vídeos e os resultados são utilizados para treinar a rede neural, a qual encontrará uma relação entre parâmetros que estão diretamente ligados a QoE e o próprio QoE. Assim, depois que a rede neural estiver treinada, o PSQA é capaz de estimar QoE em tempo real por meio de um conjunto de parâmetros. No entanto, durante a fase de implementação do PSQA é necessário seguir uma metodologia composta por três etapas, são elas:

1. Eleição de parâmetros que impactam a qualidade dos vídeos
2. Testes subjetivos de qualidade

3. Treinamento e teste da rede neural

A primeira etapa do PSQA consiste em escolher um conjunto de parâmetros que distorcem a qualidade do vídeo. Para eleição de tais parâmetros podem ser levados em consideração características de *codec*, da rede, dos vídeos (conteúdo) etc. Por isso, é nesta fase que são escolhidos e codificados um conjunto de vídeos e várias cópias de tais vídeos são produzidas para serem transmitidas pela rede. Para que os vídeos sejam transmitidos com diferentes níveis de degradação, a rede deve experimentar diferentes porcentagens de congestionamento, perda, atraso, *jitter* e largura de banda etc.

Na segunda fase cada vídeo transmitido é submetido às avaliações subjetivas, onde um conjunto de avaliadores assistem aos vídeos e são questionados para fornecerem o nível de qualidade percebida. Como foi abordado no Capítulo 3, existem normas rígidas que devem ser seguidas para a realização dos testes subjetivos.

A terceira etapa consiste em treinar uma rede neural para mapear os parâmetros que influenciam a qualidade do vídeo (selecionados na etapa 1) em qualidade percebida pelo usuário (etapa 2).

A primeira proposta do PSQA, feita em [Mohamed and Rubino 2002], foi considerada um trabalho pioneiro pois estimava QoE levando em consideração parâmetros da rede e de *codec*, tais como: taxa de bit, taxa de *frame*, tipo do *frame*, taxa de perda de pacotes e tamanho da perda durante a rajada (média do número de pacotes consecutivos perdidos). Em trabalhos anteriores eram utilizados apenas no máximo dois parâmetros de rede ou *codec* e não realizavam testes subjetivos de qualidade para apresentar o desempenho de suas propostas.

Além disso, no PSQA proposto em [Mohamed and Rubino 2002] houve a utilização do método DS (*Double-Stimulus*) para realizar os testes subjetivos de 96 vídeos distorcidos com 20 avaliadores. A rede neural empregada foi um tipo especial, denominada *Ramdon Neural Network* (RNN), a qual foi construída com uma arquitetura do tipo *feedforward* de três camadas: 5 neurônios de entrada, quatro neurônios na camada oculta e 1 neurônio de saída. Tal configuração de rede neural foi estabelecida após comparar o desempenho de redes neurais artificiais (*Artificial Neural Network* - ANN) e RNN.

O trabalho proposto em [Mohamed and Rubino 2002] serviu de base para muitos trabalhos, os quais propuseram extensões ao PSQA inicial em vários contextos. A seguir, serão apresentados tais trabalhos, assim como quais configurações foram modificadas.

4.1.1 Implementações do PSQA

Os trabalhos que implementam o PSQA podem ser divididos em quatro contextos: redes *Peer-to-Peer* (P2P), redes IEEE 802.11, transmissão de vídeo por múltiplos caminhos e transmissão de vídeo utilizando técnicas de *Scalable Video Coding* (SVC).

4.1.1.1 Implementações em redes P2P

Há pelo menos cinco trabalhos que implementam o PSQA no contexto de redes P2P, são eles: [Cancela et al. 2007], [Rodríguez-Bocca et al. 2007], [Martinez et al. 2008], [Cancela et al. 2008] e [da Silva et al. 2008].

Em [Rodríguez-Bocca et al. 2007] e [Cancela et al. 2007] o PSQA implementado leva em consideração apenas dois parâmetros que afetam a qualidade do vídeo: taxa de perda de *frames* e a média do tamanho de perda da rajada. Nestes trabalhos foram utilizados 100 vídeos distorcidos e 5 avaliadores especialistas. Os trabalhos [Martinez et al. 2008] e [Cancela et al. 2008] estenderam [Cancela et al. 2007] e [Rodríguez-Bocca et al. 2007] pois implementam um mecanismo de escolha de *peers* baseado no nível de QoE estimado pelo PSQA.

Ainda no contexto das redes P2P, [da Silva et al. 2008] propõe uma versão do PSQA diferente das versões apresentadas em [Cancela et al. 2007], [Martinez et al. 2008], [Rodríguez-Bocca et al. 2007] e [Cancela et al. 2008] pois utiliza como parâmetros de entrada: a taxa de perdas dos *frames* I, P e B. Além disso, a base de vídeos distorcidos é formada por 204 vídeos e foram utilizados apenas 10 avaliadores.

4.1.1.2 Implementações em redes IEEE 802.11

No contexto de redes IEEE 802.11, [Piamrat et al. 2009b] é o principal trabalho que implementa PSQA pois o foco deste trabalho é unicamente descrever o processo de implementação do PSQA para redes IEEE 802.11. Neste trabalho as entradas do PSQA são: taxa de perda e tamanho da perda durante a rajada. O total de avaliadores não é descrito e os testes são feitos com apenas um único vídeo. As perdas dos vídeos estão entre 0 a 10 % enquanto a tamanho de perda durante a rajada é de 1 a 3 pacotes. Além disso, os resultados apresentados em [Piamrat et al. 2009b] apresentaram desempenho superior do PSQA comparado apenas com o PSNR.

Os demais trabalhos no contexto de redes IEEE 802.11 são [Piamrat et al. 2008a], [Piamrat et al. 2008b] e [Piamrat et al. 2009a], porém tais trabalhos não fornecem descrições sobre a implementação do PSQA, uma vez que o foco destes trabalhos é utilizar o PSQA apenas como uma ferramenta de auxílio aos mecanismos propostos. Assim, a proposta geral de tais trabalhos são:

- [Piamrat et al. 2008a] propõe um mecanismo que utiliza PSQA no controle de admissão de tráfego multimídia em redes IEEE 802.11b;
- [Piamrat et al. 2008b] apresenta um mecanismo de seleção de redes sem fio fazendo o uso do PSQA e do padrão IEEE 802.11k;
- [Piamrat et al. 2009a] apresenta um mecanismo de transmissão de vídeo que ajusta o *bitrate* de transmissão de acordo com o nível de QoE estimado pelo PSQA.

Nestas três propostas considera-se sempre que há uma implementação do PSQA em cada *Access Point* (AP) da rede e o PSQA é capaz de estimar o nível de QoE de todos os clientes da rede periodicamente ou toda vez que uma nova conexão é requerida.

No entanto, muitas informações sobre a metodologia do PSQA não são apresentadas, tais como: o número de avaliadores utilizados, a quantidade de vídeos originais e degradados, a resolução dos vídeos, o nível de perda da rede etc. As únicas informações reveladas sobre o PSQA nestes trabalhos foram os parâmetros de entrada da rede neural, os quais, em [Piamrat et al. 2008b] e [Piamrat et al. 2008a] foram a taxa de perda e tamanho da perda durante a rajada, e em [Piamrat et al. 2009a] foram a taxa de perda dos *frames* I, P, B e a média do tamanho de perda da rajada do *frame* I.

4.1.1.3 Utilização de múltiplos caminhos para transmissão de vídeo

No contexto de transmissão de vídeo através de *multipath* (múltiplos caminhos) existem, no mínimo, quatro trabalhos que utilizam o PSQA para estimar QoE em tempo real, são eles: [Ghareeb et al. 2009], [Ghareeb and Viho 2010b], [Ghareeb and Viho 2010a] e [Cherif et al. 2011].

[Ghareeb et al. 2009] propõe um mecanismo de transmissão de vídeo através de múltiplos sub-fluxos e seleção de rotas que melhor se adaptam aos sub-fluxos gerados. Quando um vídeo é requisitado ao servidor de vídeo, o mecanismo proposto escolhe uma estratégia de divisão do fluxo de vídeo em sub-fluxos e escolhe rotas para tais sub-fluxos. Do lado cliente, o PSQA avalia e informa aos nós da rede o nível de QoE do vídeo transmitido.

[Ghareeb and Viho 2010b] e [Ghareeb and Viho 2010a] apresentam um mecanismo de transmissão de vídeo usando *Multiple Description Coding* (MDC). Ao requisitar um vídeo ao servidor, o cliente informa sua capacidade e velocidade para reproduzir o vídeo. Ao receber a requisição, o servidor estima a banda disponível em cada rota da rede e escolhe aquelas rotas com melhor banda disponível. Quando o cliente começa a receber o vídeo, o PSQA inicia o monitoramento de QoE e periodicamente informa ao servidor o nível de QoE do vídeo. Caso o servidor detecte que o cliente está recebendo o vídeo com qualidade baixa, então são escolhidas novas rotas para transmissão do vídeo.

No entanto, os trabalhos [Ghareeb and Viho 2010b], [Ghareeb and Viho 2010a] e [Ghareeb et al. 2009] apresentam informações muito resumidas sobre a implementação, funcionamento e resultados do PSQA. As informações explícitas em [Ghareeb et al. 2009] foram os parâmetros de entrada utilizados (taxa de perda e tamanho da perda durante a rajada), o número de avaliadores utilizados (27) e a quantidade de vídeos originais utilizados (apenas 1).

Já nos trabalhos [Ghareeb and Viho 2010b] e [Ghareeb and Viho 2010a] utilizaram 5 parâmetros como entrada: taxa de perda dos *frames* I, P, B, tamanho do GoP e MDC. O nível de perdas enfrentado pelos fluxos de vídeo está entre 0-50% para *frames* do tipo I e P, sendo que o *frame* B experimentou perdas no intervalo 0 - 100%. Além disso,

o tamanho de GoP utilizado foi de 15 e 30. As avaliações subjetivas foram realizadas com 15 pessoas e apenas um único vídeo foi utilizado.

Por fim, [Cherif et al. 2011] é o trabalho onde são apresentadas mais informações sobre a implementação do PSQA, visto que o foco deste trabalho é apresentar uma extensão do PSQA denominada ALICANTE *Pseudo Subjective Quality Assessment* (APSQA). Para a implementação dessa versão do PSQA é utilizada uma base de vídeos originais composta por 5 vídeos e uma base de vídeos distorcidos criada a partir das transmissões dos vídeos originais em uma rede cabeada onde os erros foram inseridos aleatoriamente. Apesar do APSQA utilizar um conjunto de vídeos com diferentes níveis de atividade espaço-temporal não há nenhum mecanismo de monitoramento de tais parâmetros.

As entradas que o APSQA utiliza são: taxa de perda de *frames* I, P, B e o tamanho de perda durante a rajada de cada tipo de *frame*. Além disso, os resultados apresentam uma comparação entre dois métodos *No-Reference* e dois métodos *Full-Reference*: bSoft, Demokritos, SSIM e PSNR, respectivamente. Porém os vídeos são transmitidos em um cenário cabeado.

4.1.1.4 Utilização de técnicas de SVC para transmissão de vídeo

As técnicas de SVC consistem em possuir separadamente várias codificações de um vídeo, denominados *layers*. Cada *layer* pode ser transmitida utilizando baixa largura de banda e quanto mais *layers* recebidas pelo cliente maior é a qualidade do vídeo.

Neste contexto, em [Ghareeb et al. 2011] há uma proposta para tornar o uso de SVC mais eficiente, evitando o uso desnecessário dos recursos de rede. Tal proposta consiste em medir as larguras de bandas disponíveis em todas as possíveis rotas quando um cliente requisita um vídeo ao servidor. Além disso, o cliente envia no momento da requisição um conjunto de informações sobre sua capacidade, por exemplo: resolução suportada, processamento e nível mínimo de qualidade aceitável.

Baseado nestas informações o servidor decide se o vídeo pode ser transmitido ou não, e caso a transmissão seja possível então o servidor calcula a quantidade de *layers* que o cliente tem capacidade de receber com qualidade. Assim que o cliente inicia o recebimento do *streaming* de vídeo, o PSQA começa a inferir o nível de QoE da aplicação de vídeo e envia esta avaliação para o servidor periodicamente. Quando a qualidade é inferior à qualidade definida pelo usuário como aceitável, então é feito um novo cálculo das rotas da rede.

Outro trabalho no contexto de SVC é [Singh et al. 2011]. Neste trabalho há uma proposta de extensão do PSQA para funcionar para transmissões SVC. O primeiro parâmetro a ser considerado por esta versão do PSQA é o *Instantaneous Decoder Refresh* (IDR), o qual possui definição similar ao conceito de GoP na codificação MPEG. Além disso, outro parâmetro é considerado: a porcentagem de perda da *layer* base, a porcentagem da primeira e a porcentagem da segunda camada.

O principal problema apresentado pelo trabalho [Ghareeb et al. 2011] consiste

em não fornecer nenhuma descrição de quais são os parâmetros, funcionamento e implementação do PSQA. Adicionalmente a isto, o trabalho [Singh et al. 2011] revela que os testes foram feitos com 5 vídeos codificados em H.264, não revelando quantas pessoas foram utilizadas durante a fase de testes subjetivos, nem as condições de rede utilizadas durante as transmissões dos vídeos.

4.1.2 Desvantagens do PSQA

Como foi abordado na Seção anterior, existem várias implementações do PSQA em diferentes contextos de redes. Por isso, nesta subseção objetivou-se destacar quais os pontos desejáveis para um mecanismo de estimativa de QoE em tempo real e quais desses pontos não estão presentes nas implementações do PSQA.

- O PSQA não utiliza, em nenhuma de suas implementações, um parâmetro que revele os níveis de atividade espaço-temporal dos vídeos.
- Apenas os trabalhos [Ghareeb and Viho 2010b] e [Ghareeb and Viho 2010a] utilizam o tamanho do GoP como parâmetro de entrada para a estimativa de QoE. Sabe-se que o tamanho do GoP é um parâmetro que define a quantidade de pacotes de *frames* I que trafegam pela rede.
- Nos trabalhos do PSQA, a base de vídeos distorcidos não é gerada por meio de transmissão de vídeo em condições de rede realística, pois as perdas são inseridas nos vídeos de modo aleatório. Além disso, em muitos trabalhos tais perdas são restritas a taxas entre 0 a 10 %.
- A base de vídeos originais, na maioria dos trabalhos, é composta por apenas um único vídeo e isto não é uma característica desejável para uma ferramenta de estimativa de QoE, visto que os vídeos transmitidos pelos usuários possuem diversos níveis de atividade espaço-temporal.
- As avaliações subjetivas utilizam baixo número de avaliadores ou não revelam o número de avaliadores utilizados durante os testes subjetivos. Adicionalmente a isto, muitas características não são descritas, tal como o tipo de metodologia utilizada para realizar os testes subjetivos.

4.2 *Discrete Evaluation Quality Assessment (DEQA)*

O objetivo dos trabalhos [Mu et al. 2010] e [Mu et al. 2011] é propor um *framework* chamado *Discrete Evaluation Quality Assessment* (DEQA) capaz de avaliar o nível de qualidade de serviços de IPTV (*Internet Protocol TV*). DEQA atua reconhecendo individualmente cada falha de vídeo percebida pelo usuário e mensura o impacto de tais falhas na qualidade geral do vídeo. Dessa forma a qualidade final do vídeo é dada

em função dos eventos de falha e toda vez que houver uma falha percebida pelo usuário, o DEQA armazena qual a causa da falha ocorrida seja em nível de rede ou *codec*.

Para mensurar o nível de impacto da perda ocorrida, são utilizadas funções de inspeção de pacotes que analisam diversos parâmetros, como: número de *slices*¹ presentes em cada *frame*, o tamanho do *slice*, a posição do pacote no GoP, o tamanho do *frame*, o tamanho do *buffer*, a resolução do vídeo e parâmetros de quantização. Esta análise de impacto é fornecida sem decodificar os pacotes pois alguns vídeos podem estar criptografados.

Além das falhas detectadas pela inspeção de pacotes, são feitas duas outras análises: conteúdo e erro. A análise de conteúdo é feita através de duas variáveis que medem o nível das atividades espacial temporal. A análise de erro é realizada através da quantidade de macroblocos e *frames* afetados.

Os testes subjetivos foram feitos com 20 vídeos com uma resolução 688 x 480 *pixels* codificados em H.264. O banco de vídeos distorcidos é composto por 600 vídeos e 60 pessoas foram utilizadas para realizar as avaliações subjetivas.

Podem ser observados dois problemas principais no *framework* DEQA. O primeiro está no fato de utilizar uma abordagem que exige alto poder computacional durante a transmissão do vídeo, uma vez que várias informações devem ser encontradas no cabeçalho de cada pacote. O segundo problema está no seguinte fato: para que o *framework* DEQA utilize mais informações, será necessário o acréscimo de dados nos cabeçalhos. Isto causa mais *overhead* na rede.

4.3 MintMOS

O trabalho [Venkataraman and Chatterjee 2011] utiliza uma máquina de inferência para realizar a estimativa de QoE, por meio de três parâmetros de entrada: perda, atraso e taxa de bits de codificação. A partir da variação desses parâmetros, há a construção de um conjunto de vídeos com diferentes níveis de QoE e ao conjunto de tais vídeos dá-se o nome de espaço QoE. É importante destacar que para cada vídeo no espaço QoE existe uma nota de qualidade dada por avaliadores humanos.

Quando um vídeo que não está presente no espaço QoE necessita receber a estimativa de QoE pelo MintMOS, a máquina de inferência recebe os parâmetros de entrada e produz a estimativa de QoE como saída, encontrando no espaço QoE um vídeo que possua a menor distância Euclidiana². Portanto, o vídeo escolhido é aquele que possui parâmetros de entrada com maior similaridades comparado ao vídeo avaliado e assim o vídeo de entrada recebe a mesma estimativa de QoE que o vídeo similar no espaço QoE.

¹*Slices* são agrupamentos de macroblocos, os quais devem sempre representar informações horizontais seguindo o sentido esquerda-direita de um *frame*

²Distância Euclidiana d entre dois pontos (p, q) em um plano de n dimensões é dado por:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

Para a criação do espaço QoE, foram gerados 27 vídeos com diferentes níveis de perdas, atraso e taxa de codificação (bits) do vídeo *foreman*. Os 27 vídeos foram avaliados por 80 pessoas e para testar o desempenho do MintMOS, foram utilizados 13 vídeos degradados, os quais foram criados a partir dos vídeos: *foreman*, *news* e *coastguard*. Para a avaliação subjetiva dos 13 vídeos foram utilizadas 49 pessoas.

É possível observar que para construir uma ferramenta capaz de estimar QoE de vídeos, uma base de apenas 27 vídeos não é suficiente, visto que em ambientes próximo ao que ocorre na Internet há uma variação grande de fatores que influenciam a qualidade do vídeo. Além disso, as perdas ocasionadas pela rede podem impactar de forma diferentes vídeos com características diferentes. Essa diferenciação é ocasionada pelo modo de funcionamento do *codec* e o MintMOS não utiliza nenhum parâmetro de *codec* para diferenciar tais níveis de impactos.

Além disso, por utilizar a técnica de distância Euclidiana, há uma perda de acurácia à medida que são utilizados mais parâmetros no espaço QoE, pois os pontos tendem a ficar mais distantes. Por isso, a distância Euclidiana é geralmente utilizada com poucos parâmetros, porém usar poucos parâmetros na estimativa de QoE implica em não capturar vários fatores que afetam a percepção do usuário.

4.4 *Perceived QoS (PQoS)*

Em [Koumaras et al. 2010] há a proposta de um *framework* capaz de mapear parâmetros de QoS em níveis de degradação de vídeo. Além disso, este trabalho propõe um mecanismo capaz de especificar como deve ser feita a codificação dos vídeos para serem assistidos pelos usuários com uma qualidade aceitável.

Os parâmetros de entrada deste *framework* são: taxa de bits utilizada no processo de codificação e a taxa de perda da rede. Porém a taxa de perda da rede é transformada em taxa de *frames* I, P e B decodificados. Portanto é através da taxa de *frames* decodificados e do nível de degradação ocasionada pela codificação que se estima o QoE, sendo esta estimativa feita por uma função matemática.

Porém, existem alguns pontos falhos, listados a seguir:

- as perdas utilizadas durante o processo de teste e validação da proposta são constantes e uniformes durante a transmissão do vídeo. Isto leva a proposta a ser testada em um cenário irreal.
- não há variação no tamanho GoP (12).
- assume-se que não haverá perdas de pacotes de controle do *framework*.
- não há parâmetros que torne a proposta apta a estimar QoE para vídeos com vários níveis de atividade espaço-temporal em uma rede susceptível a perdas.

- considera-se que quando um pacote de um *frame* em particular é perdido, o *frame* todo é descartado, gerando uma alta taxa de descarte de *frames* e limitando uma taxa máxima de perda na rede em 10 %.

4.5 Classificação de conteúdo de vídeo

Dentre os trabalhos que propõem formas de classificar o conteúdo do vídeo, existem três trabalhos principais: [Khan et al. 2010], [Khan et al. 2009a] e [Khan et al. 2009b]. Tais trabalhos vão além de apresentar um classificador de conteúdo pois apresentam a relação existente entre a qualidade percebida pelo usuário e o conteúdo apresentado pelo vídeo.

Para definir a relação entre conteúdo do vídeo e o QoE, há uma fase (antes da transmissão) em que os vídeos são classificados em grupos, de acordo com características de atividade espaço-temporal. As características de atividade temporal são mensuradas por meio da soma das diferenças absolutas (*Sum of Absolute Difference- SAD*) dos *frames* do vídeo. Uma das possíveis técnicas para calcular SAD consiste em obter uma média da soma da diferença entre os *pixels* dos *frames* sucessivos. O nível de atividade espacial é encontrado por meio de algoritmos mais complexos que medem a quantidade de contornos e brilho dos *frames* do vídeo.

Após extrair as características que revelam o nível de atividade espaço-temporal do vídeo, utiliza-se um classificador baseado na distância Euclidiana para formar *clusters* que possuam características de conteúdo similares. Ao todo são utilizados 12 vídeos que compõem 3 *clusters*, denominados: *Slight Movement* (SM), *Gentle Walking* (GW) e *Rapid Movement* (RM). O primeiro grupo de vídeos (SM) caracteriza-se em possuir pequena região de interesse (face) e um fundo estático. O segundo grupo de vídeos (GW) caracteriza-se por apresentar conteúdo com ângulo mais amplo e maior nível de movimentação. O terceiro grupo de vídeos (RM) é composto por vídeos de esporte e/ou com movimentação mais intensa.

Como citado anteriormente, estes trabalhos propõem um mecanismo de estimativa de QoE. Tal mecanismo é dado em função da taxa de *frames* por segundos, a taxa de envio do vídeo e a taxa de erro da rede. A função utilizada para fornecer a estimativa de QoE é adaptável individualmente para cada tipo de *cluster* por meio de coeficientes obtidos por técnicas de regressão linear. Para chegar a tal função foram realizados transmissões de vídeos em uma rede sem fio.

Os experimentos foram feitos por meio de simulação (NS-2) [NS2 2012] integrada com *Evalvid* [Klaue et al. 2003] em uma rede sem fio do tipo estruturada no padrão IEEE 802.11b. A topologia da rede é composta por dois servidores de conteúdo interligados por enlaces cabeados e um AP sem fio, responsável pela comunicação com um cliente que faz as requisições de vídeo. A taxa de perda de pacotes foi estabelecida para variar em um intervalo de 1 a 20 %, ocorrendo em 0.05 segundos e a resolução dos vídeos é de 176 x 144 *pixels*.

Estes trabalhos possuem alguns pontos falhos, uma vez que não foram realizados testes subjetivos de QoE, portanto a qualidade dos vídeos foram inferidas por meio do PSNR. Outro problema detectado em tais trabalhos foi classificar o nível de atividade espaço-temporal de vídeos baseado na técnica SAD para todos os *frames* e em técnicas de detecção de bordas para todos os *pixels* do *frame*. Por meio de tais técnicas consideram-se algumas características incompatíveis com o modo que o MPEG realiza compressão, visto que o MPEG não aplica a compressão temporal para *frames* do tipo I e a compressão espaço-temporal presentes nos *frames* B e P é aplicada entre macroblocos e não entre *frames* sucessivos.

4.6 Conclusão do Capítulo

A Tabela 2 resume os trabalhos relacionados apresentados neste capítulo. É possível observar que apenas o PSQA, PQoS e [Khan et al. 2010] foram implementados ou testados em ambientes sem fio. Além disso, apenas três trabalhos realizam avaliações subjetivas: PSQA, DEQA e MintQoS.

É possível notar também que apesar de todas as propostas utilizarem uma base de vídeos com diferentes tipo de conteúdo, apenas [Khan et al. 2010] utiliza um classificador de atividade espaço-temporal.

Tabela 2: Trabalhos relacionados.

Características	PSQA	DEQA	MintQoS	PQoS	Khan et al.
Redes sem fio	sim	não	não	sim	sim
Quantidade de vídeos originais	1 a 5	20	3	6	12
Quantidade de vídeos degradados	96 a 204	600	40	-	-
Vídeos com variação de codificação	sim	não	sim	sim	não
Avaliação subjetiva	sim	sim	sim	não	não
Quantidade de avaliadores	5 a 20	60	80	-	-
Classificador de atividade espaço-temporal	não	não	não	não	sim
Perdas de pacotes	inseridas	simulada	inseridas	real	simulada

CAPÍTULO 5

MultiQoE

Este capítulo apresenta um mecanismo de estimativa de QoE em tempo real para sistemas multimídia chamado MultiQoE. As Seções 5.1 e 5.2 apresentarão uma visão geral e o funcionamento do MultiQoE, respectivamente. A Seção 5.3 detalhará a implementação do MultiQoE.

5.1 Visão geral

Como foi abordado no capítulo 4, vários métodos de estimativa de QoE focam-se em parâmetros a nível de rede tais como largura de banda, taxa de perda, atraso e *jitter*. Porém somente os parâmetros de rede são inadequados para refletir a percepção subjetiva do usuário. Assim, parâmetros de *codec* também devem ser considerados em soluções de estimativa de QoE.

MultiQoE é um mecanismo capaz de estimar QoE de aplicações multimídia em tempo real utilizando redes neurais artificiais para prover resultados próximos à percepção humana. Como apresenta a Fig. 8, o MultiQoE utiliza redes neurais para mapear parâmetros objetivos de rede e *codec* em aspectos da avaliação subjetiva (MOS). Maiores detalhes sobre as múltiplas redes neurais será descrito em 5.3.6.

Os parâmetros de entrada do MultiQoE são denominados objetivos porque podem ser monitorados através de mecanismos de inspeção de pacotes/*codec*. Por outro lado, a saída do MultiQoE é denominado MOS estimado pois trata-se da nota subjetiva estimada. Assim, pode-se medir a acurácia do MultiQoE comparando o MOS estimado e o MOS fornecido pelos avaliadores humanos.

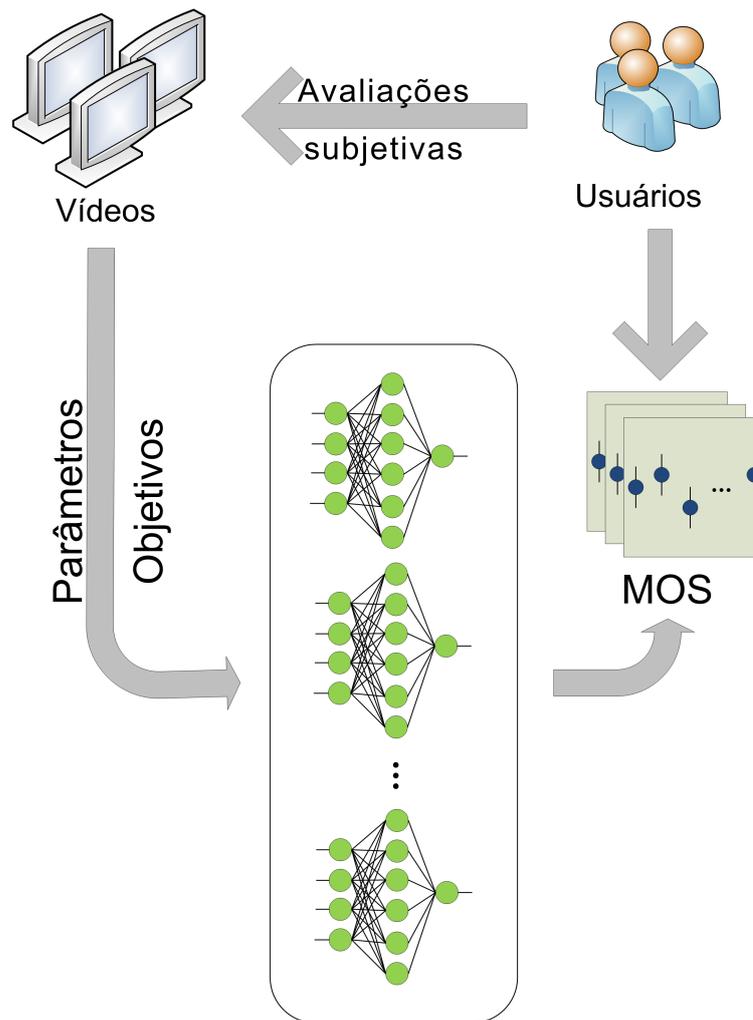


Figura 8: Visão geral do MultiQoE.

5.2 Funcionamento

Como está ilustrado na Fig. 9, o MultiQoE funciona monitorando parâmetros objetivos das aplicações de vídeo dos clientes associados aos roteadores. Abaixo estão listados os 6 parâmetros utilizados pelo MultiQoE para fornecer o MOS estimado:

1. Taxa de perda de *frames-I*
2. Taxa de perda de *frames-P*
3. Taxa de perda de *frames-B*
4. Porcentagem geral de perda de pacotes
5. Nível de atividade espaço-temporal
6. Tamanho do GoP

Para cada transmissão de vídeo iniciada em qualquer cliente da rede, há o monitoramento em tempo real de parâmetros de rede/*codec*. Com exceção do parâmetro 5, todos os parâmetros monitorados são utilizados diretamente pelo MultiQoE para fornecer o MOS estimado. No caso especial do parâmetro 5, um sistema de inferência é utilizado para fornecer ao MultiQoE o nível de atividade espaço-temporal do vídeo.

É importante ressaltar que todos os parâmetros objetivos foram escolhidos para dar ao MultiQoE a capacidade de estimar a qualidade de vídeos que possuam (i) diferentes padrões de codificação MPEG; (ii) e diferentes níveis de atividade espaço-temporal. As justificativas detalhadas para a seleção dos parâmetros objetivos do MultiQoE serão apresentadas na Seção 5.3.5.

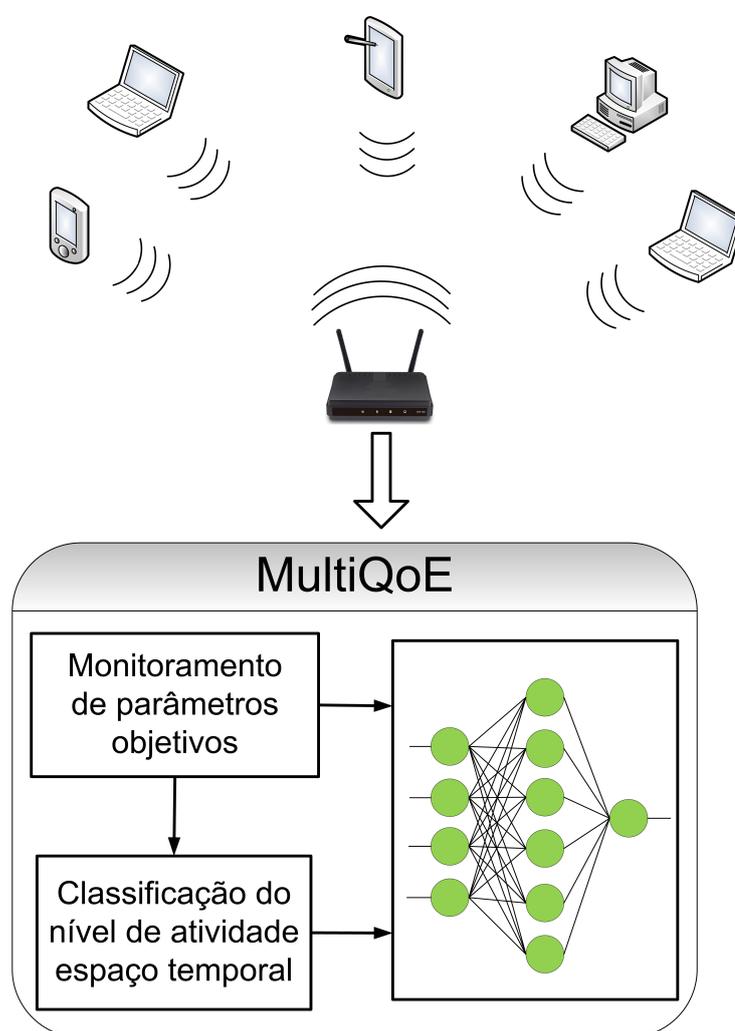


Figura 9: Funcionamento do MultiQoE.

5.3 Implementação

A metodologia utilizada para implementar MultiQoE é composta por seis etapas, como apresenta a Fig. 10 e listadas abaixo:

1. A base de vídeos originais
2. Classificação de conteúdo
3. Definição das condições de rede e composição da base de vídeos degradada
4. Avaliação subjetiva dos vídeos
5. Seleção de fatores que afetam a qualidade dos vídeos
6. Modelagem, treinamento, validação e teste das múltiplas redes neurais

A fim de fornecer maiores detalhes sobre cada etapa, esta Seção está estruturada da seguinte forma: As Subseções 5.3.1 5.3.2 e 5.3.3 apresentam a construção da base de vídeos originais, a descrição do classificador de conteúdo e as condições de rede para composição da base de vídeos degradada. Por fim, as Subseções 5.3.4, 5.3.5 e 5.3.6 descrevem os fatores de entrada escolhidos, os testes subjetivos e a modelagem das múltiplas redes neurais artificiais.

5.3.1 Base de vídeos originais

Para que o mecanismo MultiQoE seja capaz de estimar QoE de vídeos similares aos disponibilizados na Internet (diferentes níveis de atividade espaço-temporal) e como recomendado por [Seeling et al. 2007, Lotfallah et al. 2006, Seeling and Reisslein 2005], foram selecionados dez vídeos não comprimidos disponíveis em [VTL]. A Fig. 11 apresenta uma imagem representativa de cada vídeo selecionado.

Tais vídeos foram codificados em MPEG-4 com uma taxa de 25 *frames* por segundo e as durações dos vídeos são entre 10 a 12 segundos. Além disso, nenhum dos vídeos possui áudio e devido à limitação de largura de banda todos os vídeos foram codificados na resolução 352 x 288 *pixels*.

A estrutura interna do GoP foi fixada em dois *frames* B para cada *frame* P (durante toda a codificação). Para a composição final da base de vídeos originais, o tamanho do GoP foi o único parâmetro variado em cada codificação, de forma que cada um dos 10 vídeos foram codificados três vezes, cada codificação com tamanhos de GoP 10, 20 e 30, produzindo 30 vídeos.

É importante destacar que tais tamanhos de GoP foram definidos tendo em vista que os equipamentos de vídeo da indústria utilizam GoP de tamanho entre 10 e 20, no entanto como não existe valor padrão para vídeos, é comum encontrarmos vídeos com GoP 30 em ambientes que exijam maior nível de compressão devido à escassez de recursos, como é o caso da Internet.

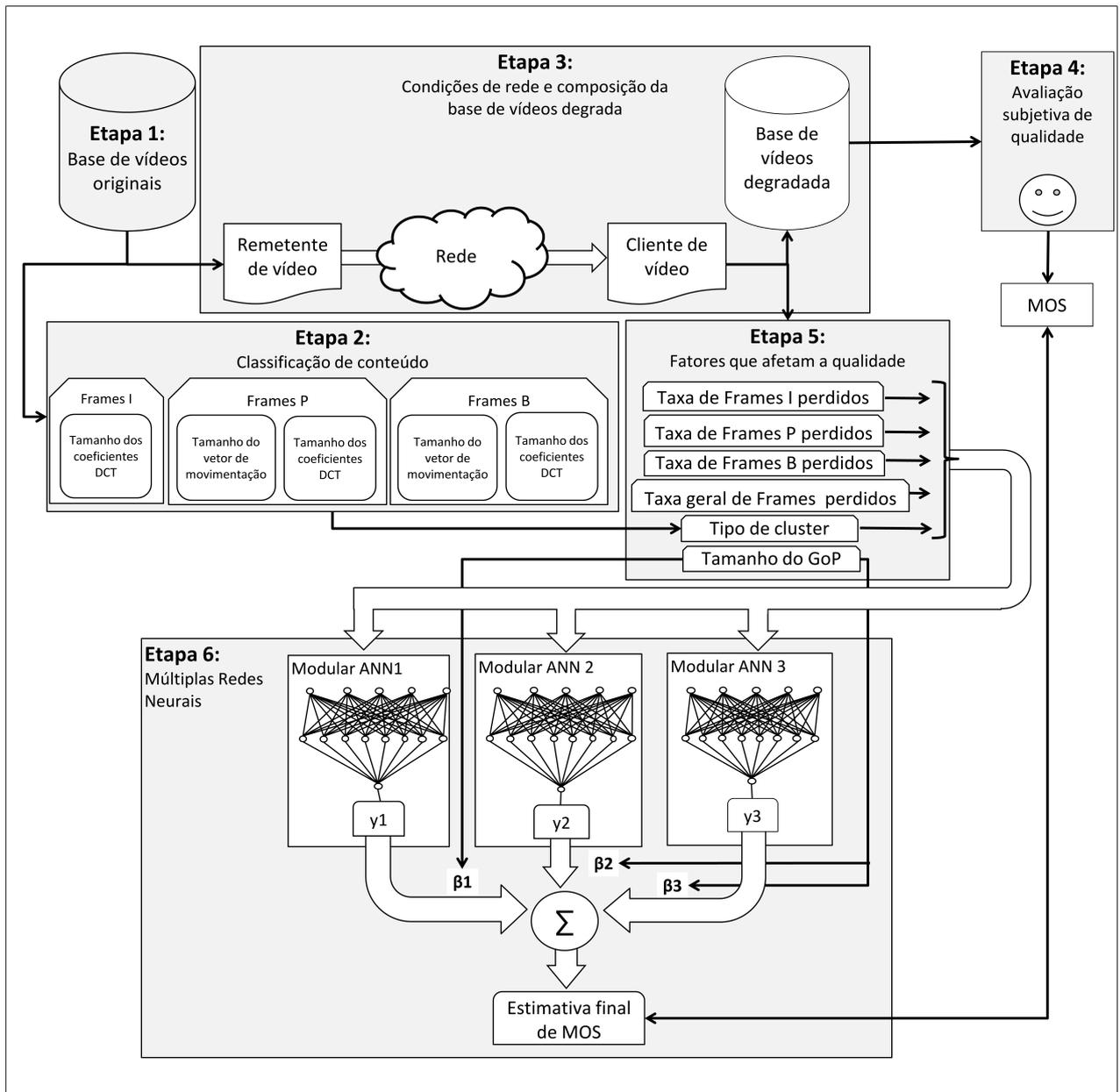


Figura 10: Etapas de implementação do MultiQoE.

5.3.2 Classificação de conteúdo

Assim como alguns trabalhos apresentam, o nível de qualidade percebido pelo usuário possui relação com a atividade espaço-temporal do vídeo [Moorthy and Bovik 2011]. Para atender a este requisito, MultiQoE utiliza um classificador de conteúdo para mensurar o nível de complexidade espaço-temporal presente no conteúdo dos vídeos. Tal classificação é realizada extraíndo o tamanho (em bits) dos coeficientes DCT (*Direct Cosine Transform*) e o tamanho do vetor de movimentação (como apresenta a Tabela 3) de cada vídeo da base de vídeos originais.

De acordo com a Tabela 3 pode-se observar que o vídeo *Flower* possui os maiores valores de coeficientes DCT, e consequentemente, o maior nível de complexidade espacial,



Figura 11: Vídeos selecionados.

Tabela 3: Tamanho médio em bits dos coeficientes DCT e dos vetores de movimentação dos vídeos

	<i>Akyio</i>	<i>News</i>	<i>Container</i>	<i>Hall</i>	<i>Mother</i>	<i>Silent</i>	<i>Mobile</i>	<i>Flower</i>	<i>Coastguard</i>	<i>Football</i>
Coeficientes DCT (<i>Frame I</i>)	80104	70802	80977	73084	67100	76399	88911	82163	42394	39682
Coeficientes DCT (<i>Frame P</i>)	9213	9792	8627	8512	8287	8429	22310	24479	14213	26302
Coeficientes DCT (<i>Frame B</i>)	1340	1894	1598	2017	988	1364	7583	7325	2820	8600
Vetor de movimentação (<i>Frame P</i>)	488	894	760	717	1132	1056	1473	1374	1744	1970
Vetor de movimentação (<i>Frame B</i>)	308	683	452	387	703	911	971	999	878	3092

enquanto o vídeo *Football* possui o maior valor de vetor de movimentação e, por isso, maior nível de atividade temporal. Dessa forma, a classificação dos vídeos em níveis de atividade espaço-temporal permite o MultiQoE ser adaptável à maioria das aplicações multimídias.

Um sistema de clusterização hierárquica, usando menor distância Euclidiana, foi desenvolvido para agrupar vídeos que possuam níveis similares de atividade espaço-temporal. A Fig. 12 apresenta a hierarquia multinível da distância Euclidiana assim como os *clusters* definidos, onde os *clusters* agrupam vídeos que possuam valores similares de

atividade espaço-temporal.

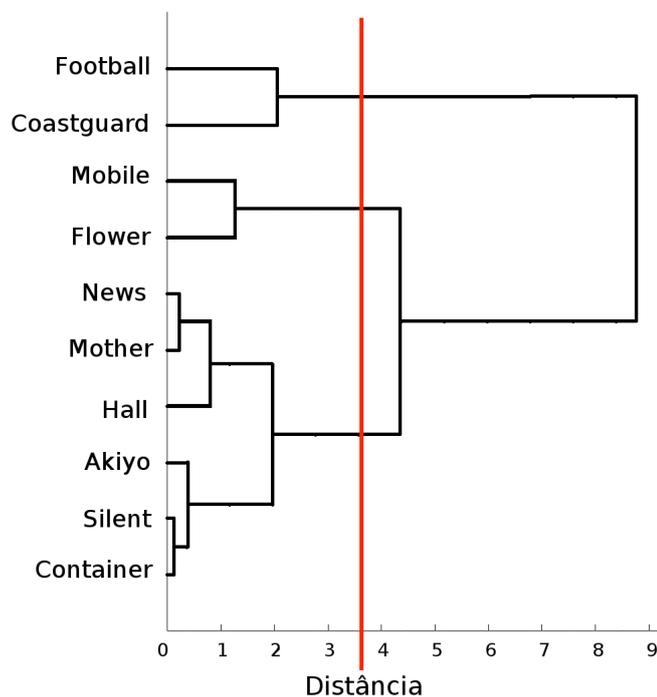


Figura 12: Árvore das distâncias Euclidianas.

De acordo com as distâncias apresentadas na Fig. 12, cada vídeo da árvore hierárquica está ligado ao vídeo ou grupo mais similar. As três maiores distâncias de ligação foram utilizadas como parâmetro para determinar a divisão dos *clusters* no conjunto de dados, conforme representa a linha vermelha na Fig. 12. Tal divisão foi necessária para produzir três *clusters* com similaridades consistentes.

Usando os três *clusters* apresentados na Fig. 12 e pela análise dos coeficientes DCT e o tamanho dos vetores de movimentação, é possível classificar os três *clusters* em: Baixo nível Espaço-Temporal (BET), Alto nível Espacial e Médio nível Temporal (AEMT) e Médio nível Espacial e Alto nível Temporal (MEAT).

Como apresenta a Tabela 4, o *cluster* BET é composto por vídeos com conteúdo caracterizado por ter pequena região de interesse e um fundo estático. O *cluster* AEMT possui vídeos com constante movimentação de câmera e cenas com muita informação visual. Por fim, o *cluster* MEAT é composto por vídeos com rápida movimentação de câmera ou rápida movimentação de fundo com médio nível de informação visual.

5.3.3 Condições de rede e composição da base de vídeos degradada

Depois de construir a base de vídeos originais com diferentes níveis de atividade espaço-temporal e com diferentes tamanhos de GoP, é necessário transmitir tais vídeos em um cenário que reflita ambientes de redes na prática. Diversos tipos de tecnologias

Tabela 4: Características e classificação dos vídeos selecionados

Tipo de <i>cluster</i>	Característica de conteúdo	Vídeos
Baixo nível Espaço-Temporal (BET)	Pequena região de interesse com fundo estático	<i>Akiyo, News, Container, Hall, Mother, Silent</i>
Alto nível Espacial e Médio nível Temporal (AEMT)	Movimento contínuo de câmera e cenas com detalhes pequenos	<i>Mobile, Flower</i>
Médio nível Espacial e Alto nível Temporal (MEAT)	Rápido movimento de câmera ou rápida mudança de fundo	<i>Coastguard, Football</i>

de transmissão poderiam ter sido escolhidas para implementar o mecanismo MultiQoE, porém a tecnologia em malha sem fio foi escolhida devido ao crescente aumento de popularidade e dos desafios existentes na transmissão de vídeo em um ambiente de múltiplos saltos.

Dessa forma, o *backbone* da rede em malha sem fio (*Wireless Mesh Network - WMN*) da UFPA (Universidade federal do Pará) foi escolhida como apresenta a Figura 13.

Para garantir que cada fluxo de vídeo não sofresse uma taxa de degradação arbitrária, um modelo de perda exponencial foi aplicado. Como apresenta a Figura 14, esse modelo garante que a maior parte dos vídeos sofreu até 10% de perda e uma menor parte dos vídeos sofreu mais de 50% de perdas.

5.3.4 Avaliação subjetiva

Com a base de vídeos degradada, um plano de teste foi cuidadosamente definido para avaliar subjetivamente todos os vídeos. As avaliações subjetivas foram realizadas questionando um grupo de avaliadores, o nível de qualidade dos vídeos degradados, utilizando as recomendações ITU [P.900 2002, ITUR-B500-11 2002].

Os avaliadores foram questionados para fornecer uma nota de qualidade com relação ao nível de impacto visual das distorções. A técnica utilizada para realizar os testes subjetivos foi *Absolute Category Rating* (ACR) especificada na recomendação [ITUR-B500-11 2002], também conhecida como *Single Stimulus* (SS) especificada em [P.900 2002].

Durante a realização dos testes subjetivos, seguindo as especificações das técnicas SS e ACR, somente o vídeo degradado (alvo da avaliação) é exibido uma única vez. A escolha do método SS/ACR foi motivada por ser mais aplicável a estudos que objetivam avaliar uma grande quantidade de vídeos, uma vez que tal técnica reduz significativamente o tempo necessário para conduzir o estudo.

O número de avaliadores envolvidos nos testes subjetivos foi, ao todo, 55. To-

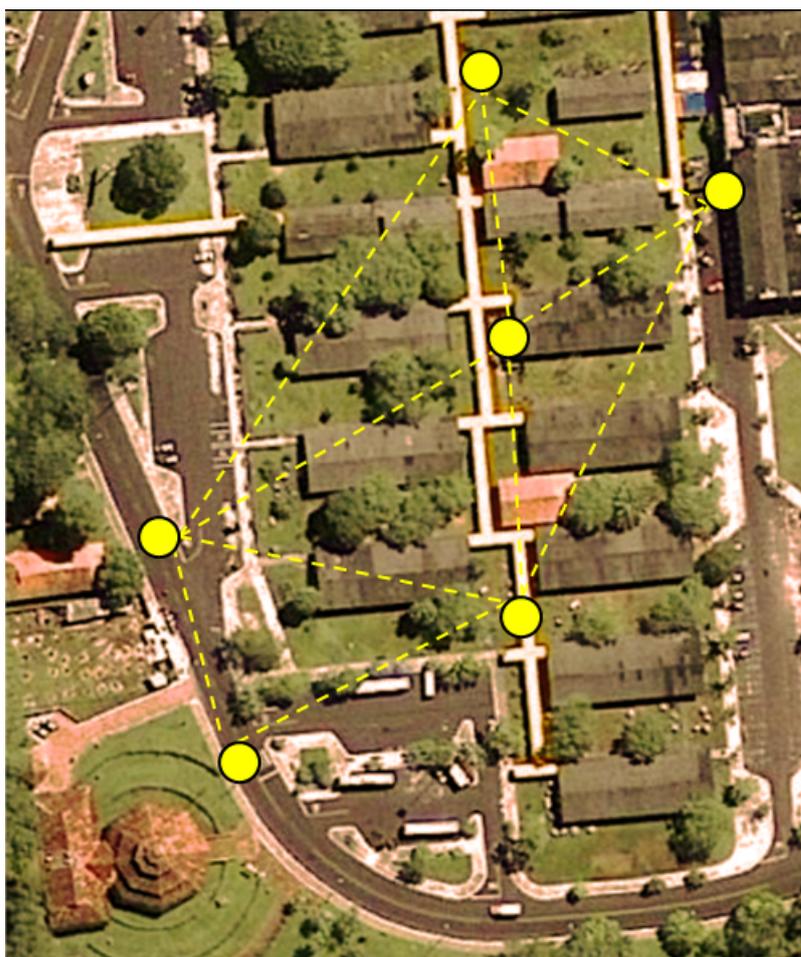


Figura 13: Rede em malha sem fio da UFPA.

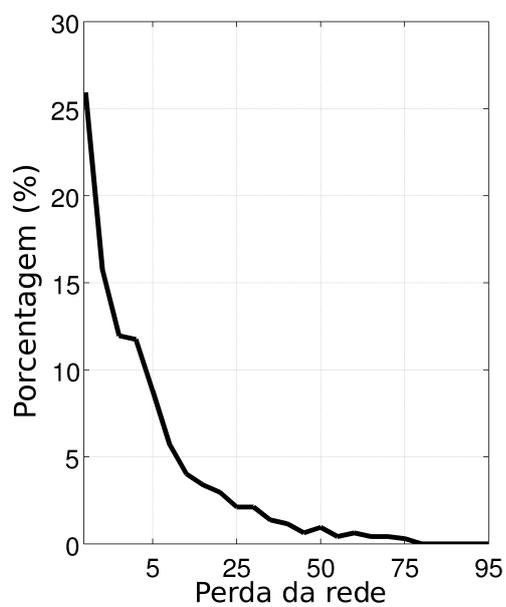


Figura 14: Taxa de perdas de pacotes em todos os experimentos.

das as pessoas possuíam visão normal e suas idades estavam entre 18 a 45 anos. Os participantes incluíam estudantes de graduação, pós-graduação e outros funcionários da universidade.

Para cada vídeo na base degradada, o cálculo da nota média, \bar{u}_{jkr} , foi obtida por:

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^N u_{ijkr} \quad (5.1)$$

onde:

u_{ijkr} : nota do avaliador i para condições de teste j , sequência/vídeo k , repetição r

N : número de avaliadores.

No passo seguinte, todas as notas médias foram associadas com um intervalo de confiança de 95%, o qual é dado por $[\bar{u}_{jkr} - \delta_{jkr}, \bar{u}_{jkr} + \delta_{jkr}]$ onde

$$\delta_{jkr} = 1.96 \frac{S_{jkr}}{\sqrt{N}} \quad (5.2)$$

Onde o desvio padrão para cada vídeo, S_{jkr} , é dado por:

$$S_{jkr} = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_{jkr} - u_{ijkr})^2}{(N-1)}} \quad (5.3)$$

A Figura 15 revela o histograma das notas (MOS) obtidas após o cálculo apresentado pela Equação 5.2. A nota 3 (regular) foi o valor mais selecionado durante a avaliação dos vídeos (35,79%), enquanto a nota 2 (pobre, 30,21%) e 4 (bom, 16,53%) foram selecionadas em segundo e terceiro lugar. Finalmente, em 15,16% dos vídeos a nota fornecida pelos avaliadores foi 1 (ruim) e em 2,32% dos vídeos a nota foi 5 (excelente).

5.3.5 Seleção de fatores que afetam a qualidade

Para que MultiQoE possa fornecer estimativa de QoE com alto grau de correlação com as avaliações subjetivas, considera-se que a taxa de perda de *frame* I, P e B são parâmetros importantes para inferir a qualidade de vídeo devido à estrutura hierárquica de codificação. Além desses parâmetros, são considerados também as características de conteúdo do vídeo (tipo de *cluster*) e tamanho do GoP. Tais parâmetros serão descritos com maiores detalhes nas subseções seguintes.

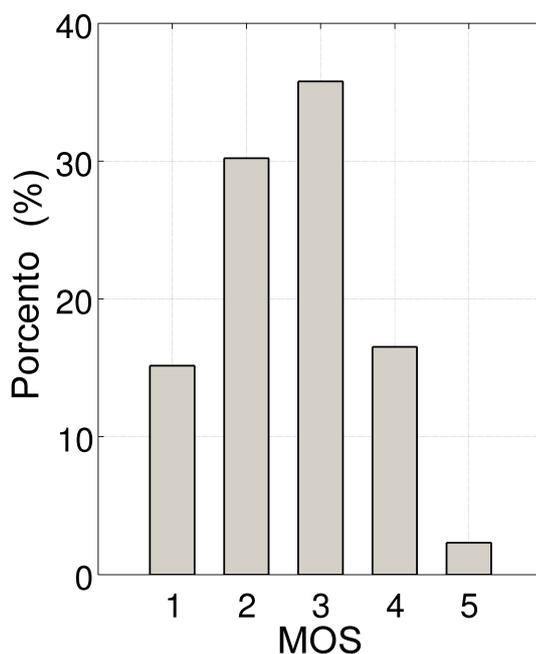


Figura 15: Histograma das notas obtidas.

5.3.5.1 Taxa de perda de *frames* I, P e B

As taxas de perdas dos *frames* I, P e B ajudam a identificar níveis de impactos distintos que um único pacote perdido durante a transmissão afeta a qualidade de experiência. Tais níveis distintos de impactos na qualidade percebida ocorrem pois durante uma transmissão de vídeo MPEG-4 existem três possibilidades de perdas, listadas abaixo:

1. se o pacote descartado é do tipo I, o erro proveniente do descarte deste pacote será propagado por todos os *frames* restantes do GoP, uma vez que o decodificador MPEG utiliza os *frames* I como referência para todos os *frames* de um GoP. Quando isto ocorre, a qualidade do vídeo voltará ao normal apenas quando o decodificador receber um *frame* I não degradado completo.
2. se o pacote descartado é do tipo P, falhas visuais serão percebidas nos *frames* posteriores ao *frame* corrompido.
3. caso o pacote descartado seja do tipo B então os danos afetarão apenas o *frame* B, uma vez que os *frames* B não são utilizados como referência para outros *frames*.

5.3.5.2 Nível de atividade espaço-temporal

Além das taxas de perdas de *frames* I, P e B, o nível de atividade espaço-temporal é um parâmetro que o MultiQoE leva em consideração, uma vez que o nível de compressão espacial e temporal do vídeo altera as probabilidades de haver perda em cada tipo de

frame. Tal probabilidade é alterada, pois quando um *frame* possui maior compressão implicará em menor número de pacotes daquele *frame* durante a transmissão do vídeo. A Tabela 5 apresenta a influência do conteúdo do vídeo na composição do fluxo de vídeo.

Tabela 5: Influência do conteúdo na composição do fluxo de vídeo.

Característica do conteúdo	Consequência na composição do fluxo de vídeo
Alto nível de atividade temporal	Aumenta o número de pacotes P e B
Alto nível de complexidade espacial	Aumenta o número de pacotes I, P e B
Baixo nível de atividade temporal	Diminui o número de pacotes P e B
Baixo nível de complexidade espacial	Diminui o número de pacotes I, P e B

No caso de vídeos com alto grau de atividade temporal haverá uma aumento no número de pacotes do tipo P e B uma vez que os *frames* P e B realizam compressão temporal. Portanto, quando há o aumento do número de pacotes I na rede pode-se concluir que trata-se de um vídeo com maior complexidade espacial, no entanto o aumento do número de pacotes P ou B não permite concluir se a causa é o nível de atividade temporal ou complexidade espacial do vídeo.

5.3.5.3 Tamanho do GoP

Outro fator utilizado pelo MultiQoE é o tamanho do GoP, o qual é determinado pelo número de *frames* entre dois *frames* I no fluxo MPEG. Para demonstrar como o tamanho de GoP exerce influência sobre a composição do fluxo MPEG, a Figura 16 apresenta a porcentagem de pacotes I, P e B que compõem os fluxos dos dez vídeos selecionados para os GoP 10, 20 e 30.



Figura 16: A influência do GoP na composição do fluxo de vídeo.

É possível observar que para o GoP 10 (Figura 16), os vídeos *Akiyo*, *Coastguard*, *Hall*, *Mother*, *News* e *Silent* são compostos em sua maior parte por pacotes do *frame* I. Os demais vídeos, que possuem menor nível de complexidade espacial, possuem menor porcentagem de pacotes I. É possível observar, também, que os vídeos codificados com GoP 20 e 30 passam a ser composto em maior parte por *frames* do tipo P. Isso ocorre basicamente devido ao aumento do GoP, o que ocasiona a redução quantidade de *frames* I no fluxo de vídeo.

Portanto, o tamanho do GoP é considerado um parâmetro essencial pelo MultiQoE, uma vez que seu tamanho é determinante para a porcentagem de pacotes I, P e B que compõem o fluxo transmitido, alterando dramaticamente a maneira como as perdas na rede afetam a qualidade de vídeo .

5.3.6 Múltiplas redes neurais artificiais

Antes de apresentar as múltiplas redes neurais (*Multiple Artificial Neural Networks* - MANNs) do MultiQoE, é necessário justificar a utilização de redes neurais para a tarefa de estimativa de qualidade de vídeo em tempo real. Portanto, estão listadas abaixo as principais características que levaram a utilização das ANNs.

- Generalização: as ANNs são capazes de infer respostas corretas mesmo com entradas desconhecidas. Esta característica é útil no contexto deste trabalho pois se todas as combinações possíveis de entrada e saída fossem necessárias, deveria haver um número inviável de avaliadores humanos.
- Auto-adaptação: em problemas em fase de estudo, como é o caso da avaliação subjetiva de vídeo, onde nem todas as relações estão claras, as ANNs são capazes de se auto-adaptar. enditemize

MANN têm sido utilizadas em muitas áreas de pesquisa para solucionar problemas que envolvem classificação e extração de características. As MANNs permitem decompor tarefas complexas em tarefas menores e especializadas pois cada ANN é treinada para se tornar especialista em algum aspecto da tarefa geral [Khashei et al. 2012].

Assim, é possível explorar a capacidade especialista e solucionar problemas que não poderiam ser solucionados efetivamente utilizando uma única ANN. Portanto, a técnica de utilizar MANNs possui melhor capacidade, alto desempenho e permite melhor modelagem [Hashem 1997].

Como concluído na Seção 5.3.5.3, o tamanho do GoP exerce grande influência na composição do fluxo de vídeo transmitido na rede e portanto é um fator determinante para o método MultiQoE e por isso foi utilizada uma ANN para cada tamanho de GoP. Assim, a tarefa de estimar QoE foi dividida em três ANNs especializadas. Cada ANN foi treinada com uma sub-base específica, formada pelos tamanhos de GoP 10, 20 e 30 (Fig. 10). Dessa forma, a saída que cada ANN fornece é especializada para um tamanho de GoP em particular e a estimativa final de QoE é obtida pela combinação de cada ANN.

Existem várias formas de combinar as três saídas para produzir o resultado final, porém foi selecionado a combinação linear de ANNs proposto em [Hashem 1997] porque tal método fornece acurácia superior comparada com uma única ANN. A técnica de combinação linear de ANNs realiza a média ponderada das saídas das ANNs. Dessa forma, tal técnica possibilita o ajuste dos pesos durante a fase de teste e treino para que a combinação das saídas das ANN's obtenha melhor desempenho. No caso do MultiQoE,

definiu-se que os pesos assumiriam valores entre 0 a 1 de acordo com o tamanho de GoP do vídeo alvo da estimativa, onde 1 indica nível máximo de especialização de uma ANN para a tarefa, enquanto 0 é aplicado para a saída de uma ANN que não possui especialização para uma dada tarefa.

Cada peso multiplica unicamente a saída de uma rede neural e quando o tamanho do GoP do vídeo avaliado é o mesmo do tamanho de GoP previamente definido para uma rede neural em particular, então a saída de tal rede recebe o valor máximo de peso. Portanto, os pesos assumem valores que são inversamente proporcionais à distância entre o GoP do vídeo avaliado e o tamanho de GoP definido previamente para cada ANN.

Cada ANN utiliza o algoritmo de treinamento *Levenberg-Marquardt* (LM) pois tal algoritmo de treinamento fornece uma solução numérica para o problema de minimização [Khashei et al. 2012] e tem sido extensivamente aplicado como método de treinamento de ANNs para atingir alto nível de convergência e rápida velocidade de treinamento como exigido pelo MultiQoE.

A estrutura proposta para cada ANN é composta de várias entradas e uma única saída. Cada ANN é do tipo *feedforward* e *backpropagation* composta por 5 neurônios de entrada na camada de entrada, 12 neurônios na camada escondida e 1 neurônio na camada de saída. Cinco variáveis foram identificadas como entradas de cada ANN: porcentagem de perdas para cada tipo de *frame* (I, P e B), porcentagem total de perda e tipo de cluster do vídeo. É importante destacar que a saída do MultiQoE é o MOS estimado, sendo que as fases de treinamento e teste da MANN foi implementado em *MATLAB* [Demuth and Beale 1993].

Depois dos observadores avaliarem cada vídeo degradado, o processo de treinamento é atingido ao obter um mapeamento entre os parâmetros de entrada e o MOS. A fase de validação da rede foi efetuada através da técnica denominada de validação cruzada afim de reduzir o erro de generalização da rede.

Uma vez que as MANNs foram treinadas e validadas, MultiQoE pode ser utilizado em tempo real para fornecer estimativa de QoE sem necessidade da interação com humanos. Apenas é preciso medir os parâmetros de entrada no tempo t e entrar com esses valores na MANN para obter a qualidade percebida instantaneamente.

5.4 Conclusão do Capítulo

Este capítulo apresentou a mecanismo de estimativa de QoE chamado MultiQoE. Este mecanismo utiliza parâmetros da codificação de vídeo e da rede para fornecer a estimativa de QoE de um vídeo em tempo real. Além disso, o MultiQoE utiliza múltiplas redes neurais para aprender os aspectos subjetivos da avaliação humana. Assim, o MultiQoE foi implementado para ser capaz de estimar a qualidade de vídeos que possuam diferentes níveis de atividades espaço-temporal, condições de rede e codificação.

CAPÍTULO 6

Estudo de caso

Este capítulo apresentará um estudo de caso, o desempenho e os benefícios do MultiQoE ao estimar a qualidade de vários vídeos em uma rede em malha sem fio. Na Seção 6.1 há descrições sobre as configurações utilizadas e a Seção 6.2 apresenta a análise dos resultados obtidos.

6.1 Configurações do ambiente utilizado

Como apresentado na Figura 17, o cenário do estudo de caso é baseado na rede em malha sem fio da UFPA. O cenário compreende uma região amazônica, com alguns prédios e áreas de estacionamento. O *backbone* da rede é composto por seis roteadores, onde dois deles possuem a função de *gateway*. Além do *backbone*, um cliente foi posicionado para receber o fluxo de vídeo proveniente do *backbone* em malha. Vale ressaltar que em cada transmissão de vídeo a localização geográfica do cliente foi determinada por uma função de distribuição uniforme.

É necessário destacar que apesar do MultiQoE ser um mecanismo de estimativa de QoE implementado para redes em malha sem fio, pode haver a adaptação deste mecanismo para outras tecnologias de transmissão.

Os experimentos foram executados por meio do *Network Simulator* (NS) 2.34 [NS2 2012] em conjunto com a ferramenta *Evalvid* [Klaue et al. 2003]. A Tabela 6 apresenta os parâmetros gerais utilizadas durante a simulação. Para o PSQA, a RNN foi utilizado *Random Neural Network Simulator 2* (RnnSim 2) [Abdelbaki 1999]. Para o MultiQoE, as múltiplas redes neurais foram construídas utilizando o *ToolBox* padrão do *MATLAB* [Demuth and Beale 1993]. As métricas objetivas foram calculadas utilizando a ferramenta VQMT MSU [Vatolin et al. 2009].



Figura 17: Composição do *Backbone* da rede em malha sem fio da UFPA.

Tabela 6: Parâmetros de simulação configurados no simulador.

Parâmetro	Valor
Frequência	2.4 Ghz
Nº de roteadores	6
Modelo	IEEE 802.11g
Tipo de antena	Omnidirectional
Ganho da antena	12 dB
Altura da antena	8 metros
Codec de vídeo	MPEG-4
Taxa de <i>frames</i>	25 <i>frames/s</i>
Path Loss Exponent	2.59
Shadowing Deviation	4.4

Através da ferramenta *Evalvid*, foi possível transmitir os fluxos reais de vídeos no cenário proposto. Para fornecer uma base de dados com número suficiente de vídeos, cada vídeo selecionado foi transmitido 90 vezes sofrendo variadas condições de rede (perda, atraso, *jitter* etc) e com diferentes tamanhos de GoP. Tais transmissões resultaram em um total de 900 vídeos, onde 810 foram selecionados randomicamente para pertencer à

base de treino e os 90 vídeos restantes formaram a base de teste.

6.2 Análise dos resultados

A Figura 18 revela que, em média, a estimativa fornecida pelo MultiQoE foi a mesma que o MOS em 90 % dos vídeos avaliados na base de teste, enquanto PSQA alcançou apenas 41%. Portanto, as vantagens do MultiQoE sobre o PSQA são claras e demonstram sua acurácia e eficiência para estimar o nível de qualidade de vídeos com diferentes tipos de conteúdo e em sistemas multimídia sem fio.

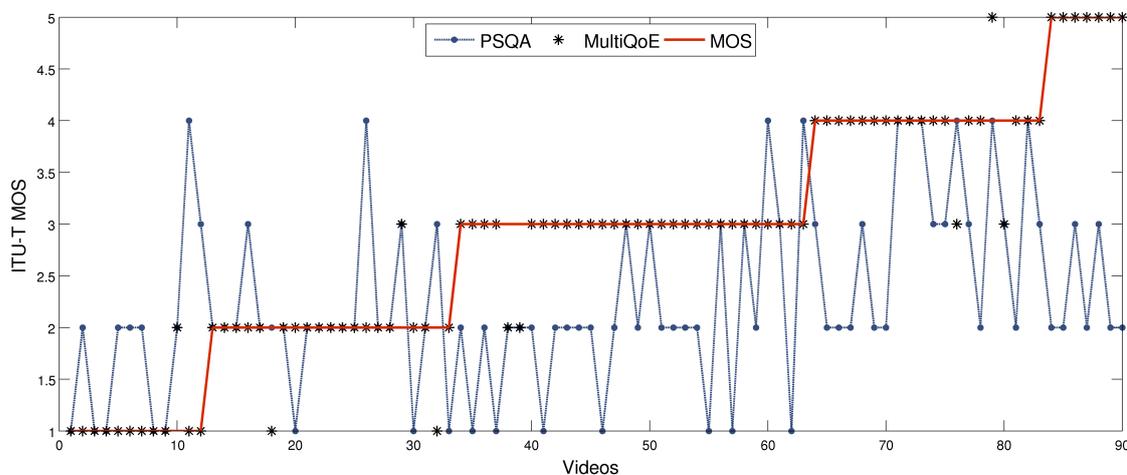


Figura 18: MultiQoE vs PSQA.

Comparado com as principais métricas objetivas, MultiQoE apresenta resultados ainda melhores para mensurar o nível de qualidade de vídeos. A Figura 19 apresenta o desempenho do MultiQoE comparado com PSNR. É possível observar que PSNR apresenta notas correlatas com o MOS em poucos casos. A estimativa de qualidade fornecida corretamente pelo PSNR foi em média de 15.55% dos vídeos avaliados. Um dos fatores que prejudica o PSNR na tarefa de estimar a qualidade de vídeo é o fato de não considerar o sistema visual humano e nem a estrutura de codificação do MPEG.

A métrica SSIM inclui em seu algoritmo de estimativa de QoE parâmetros à nível de usuário, tais como cor, brilho e informações de estrutura. Por isso SSIM tende a fornecer resultados mais exatos que o PSNR. Em média, como apresenta a Figura 20, SSIM estimou a qualidade corretamente em apenas 18.88% dos vídeos avaliados. De um modo geral, a Figura 20 revela que o SSIM não possui desempenho satisfatório para refletir a opinião do usuário para avaliar vídeos com diferentes padrões de perda, movimentação e complexidade espacial.

A métrica VQM é baseada em algoritmos que extraem informações de movimentação temporal e complexidade espacial do vídeo, além de utilizar modelos matemáticos baseados nos usuários. Portanto, VQM é considerada uma métrica objetiva mas com características subjetivas. A Figura 21 apresenta as notas obtidas pela métrica

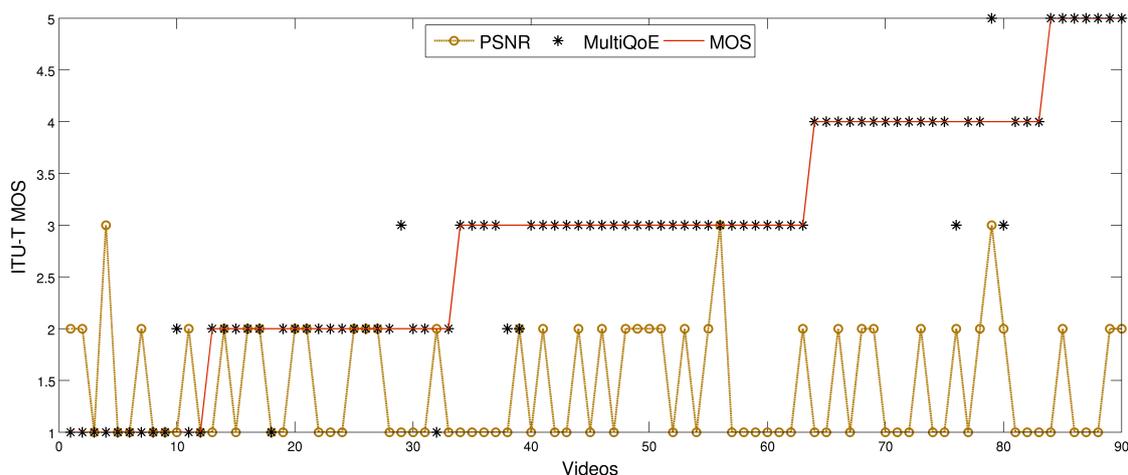


Figura 19: MultiQoE vs PSNR.

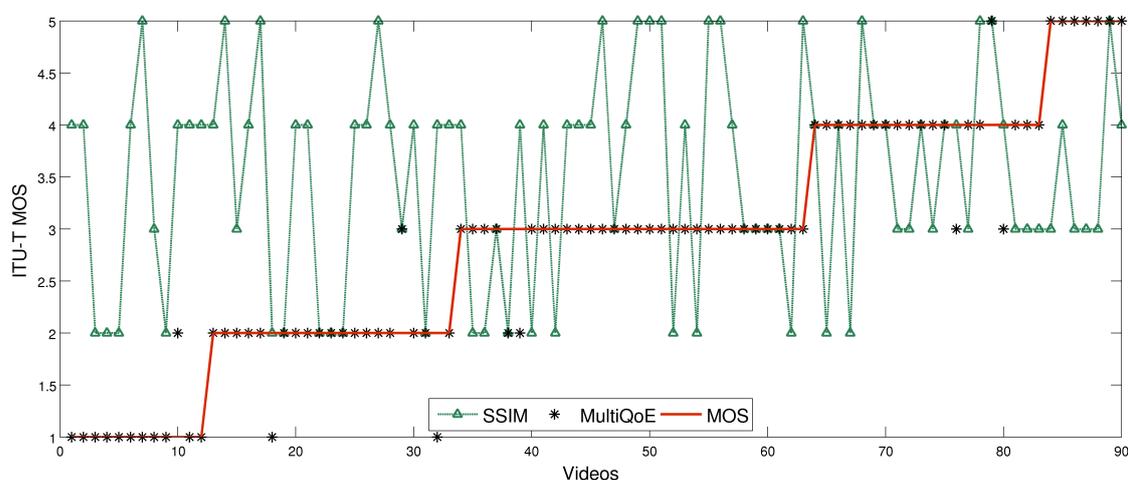


Figura 20: MultiQoE vs SSIM.

VQM. Em média VQM produziu notas de qualidade iguais às fornecidas pelos avaliadores humanos em 27.77% dos vídeos avaliados. Isto revela que dentre as métricas objetivas, VQM é a métrica com maior nível de acurácia, porém ainda não é capaz de atingir um nível de acerto alto.

Para resumir o desempenho do MultiQoE, o erro quadrático médio (*Mean Square Error* - MSE) do MultiQoE, PSQA, PSNR, SSIM e VQM para os tamanhos de GoP e para os *clusters* são apresentados nas Figuras 22 e 23, respectivamente.

A Figura 22 apresenta o MSE de cada método de estimativa de QoE para cada tipo de *cluster* de vídeo. É possível observar que para o *cluster* BET o MultiQoE possui alto desempenho com MSE de apenas 0.14, enquanto PSQA, PSNR, SSIM e VQM possuem 0.91, 2.77, 3.12 e 2.52, respectivamente. É importante destacar também que MultiQoE possui alta acurácia para os *clusters* AEMT e MEAT com MSE de 0.07 e 0.14 contra 0.64 e 0.64 do PSQA. Portanto, MultiQoE apresenta-se como uma solução para estimativa de qualidade de experiência mais estável comparada com as métricas objetivas e PSQA.

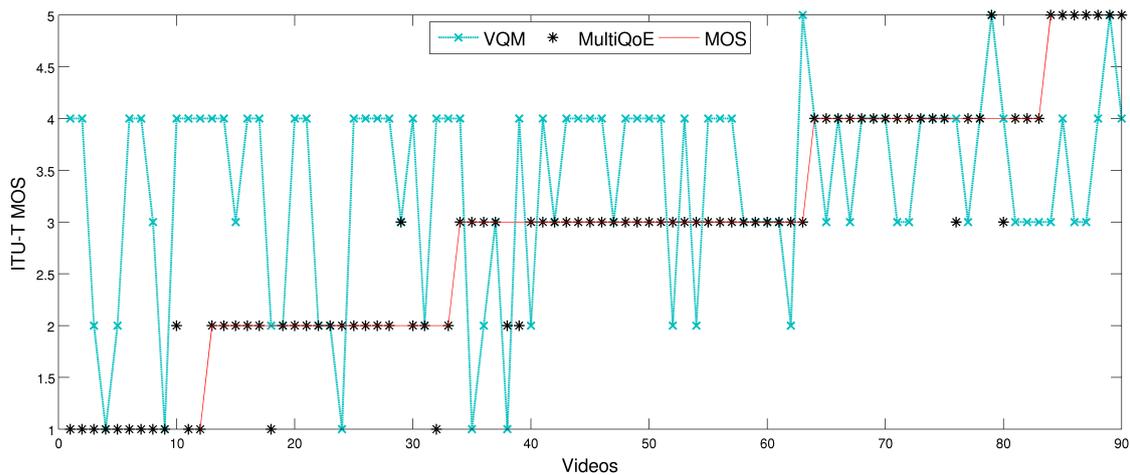


Figura 21: MultiQoE vs VQM.

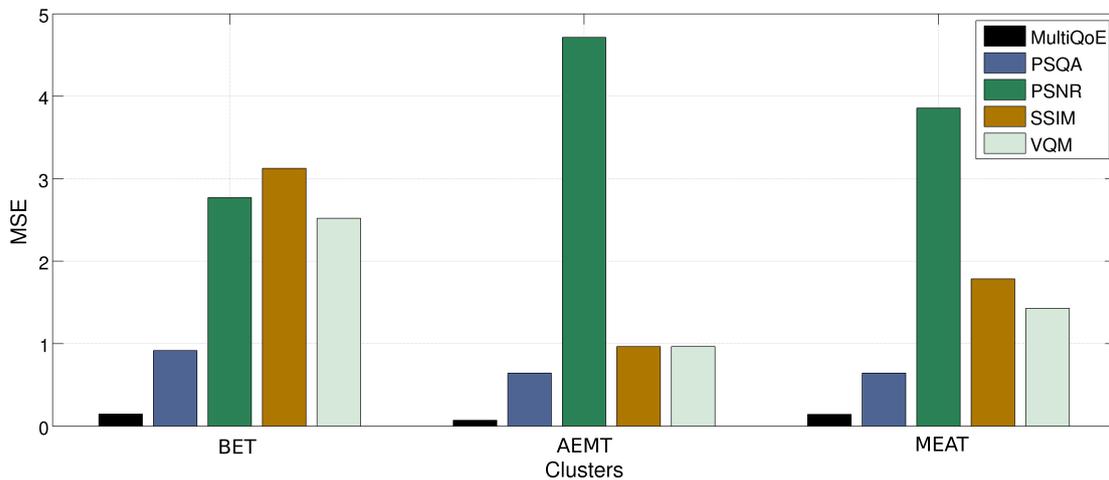


Figura 22: MSE para cada cluster.

Tendo em vista que o GoP é um fator que exerce grande influência no nível de QoE apresentado por um vídeo, a Figura 23 apresenta o MSE dos mecanismos de estimativa de QoE para cada tamanho de GoP. No caso do GoP 10, MultiQoE e PSQA apresentam MSE de 0.1 e 1.04, respectivamente. Para o GoP 20, MultiQoE apresenta um desempenho de 0.15, enquanto PSQA apenas 0.65. Finalmente, para vídeos com GoP 30, o erro produzido pelo MultiQoE é de 0.15, enquanto PSQA 0.3. É possível observar que no cenário de GoP de 30 o método PSQA apresentou uma melhora no desempenho porque neste cenário a perda (total ou parcial) de *frame* I possui maior impacto na qualidade de vídeo. Os métodos PSNR, VQM e SSIM apresentaram um desempenho insatisfatório para todos os cenários de GoP pois em todos eles o MSE ficou entre 1.6 e 5.19.

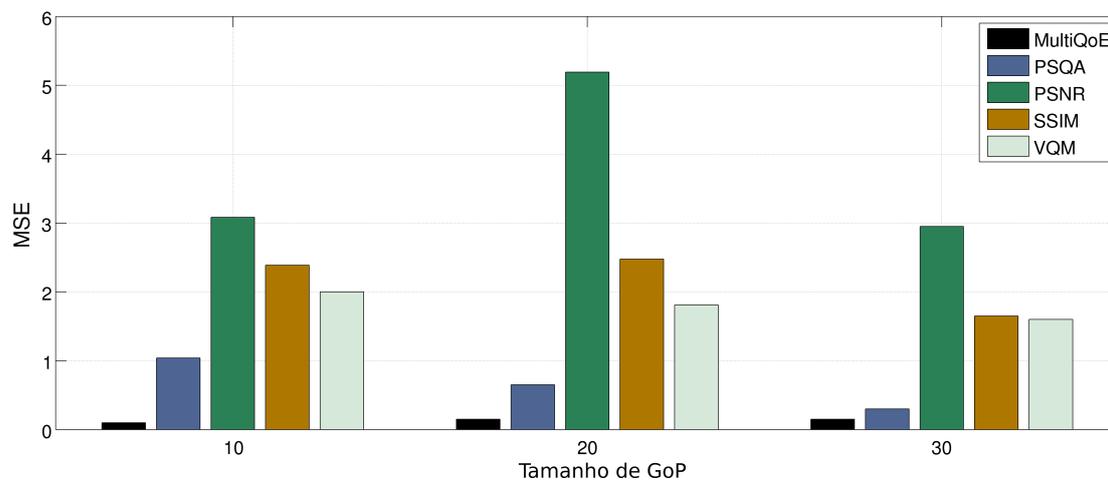


Figura 23: MSE para cada tamanho de GoP.

6.3 Conclusão do capítulo

Este capítulo apresentou que 810 vídeos (base de treino) foram utilizados para treinamento dos métodos MultiQoE e PSQA. Além disso, 90 vídeos foram utilizados em experimentos que estimavam o nível de QoE por meio das seguintes métricas e métodos: PSNR, SSIM, VQM, PSQA, MultiQoE e MOS. Os resultados apresentados revelam que o MultiQoE possui alta correlação com o MOS e melhor desempenho do que o PSQA, VQM, SSIM e PSNR.

CAPÍTULO 7

Conclusões e trabalhos futuros

7.1 Conclusões gerais

Métodos capazes de estimar com sucesso o nível de QoE são vitais para o sucesso de sistemas multimídias em cenários sem fio e tem recebido muita atenção da indústria e da academia nos últimos anos. Além disso, os atuais métodos de estimativa de QoE, baseados apenas em parâmetros de QoS, não são capazes de refletir a experiência do usuário.

É um desafio estimar em tempo real a qualidade de aplicações de vídeo, principalmente quando busca-se considerar um cenário próximo ao real. As diferentes possibilidades de codificação, os vários eventos de perdas durante a transmissão e os aspectos subjetivos da avaliação do usuário são alguns, dos muitos, fatores que podem influenciar a nota de qualidade do usuário. Porém os mecanismos de estimativa de QoE devem possuir tal capacidade e abrangência, uma vez que a Internet está repleta de vídeos com diferentes codificações e sujeita a diversos tipos e níveis de degradação.

Além disso, as redes em malha sem fio são cenários que torna o trabalho mais desafiador, pois a transmissão de vídeo em um ambiente *mesh* utiliza um *backbone* sem fio, isto possibilita os múltiplos saltos e outros fatores agravantes. Por este fato, alguns dos trabalhos relacionados propõem mecanismos para cenários em que a perda ocorre de maneira controlada (inseridas aleatoriamente) e dentre os poucos trabalhos que fazem transmissão de vídeo em redes sem fio, nenhum (dentre os pesquisados) utiliza um cenário *mesh*. Vale ressaltar que apesar do MultiQoE ter sido implementado para uma rede em malha sem fio, o mecanismo proposto não está restrito a este cenário, uma vez que outras tecnologias de transmissão podem ser utilizadas, desde que as etapas de implementação do MultiQoE, descritas no Capítulo 5, sejam seguidas.

Para modelar a percepção do usuário buscou-se utilizar a técnica de inteligência computacional denominada redes neurais. Tal técnica é capaz de modelar variáveis não triviais (ainda em estudo), como por exemplo, aspectos subjetivos da avaliação humana. Outro fator importante a ser considerado para a escolha da técnica de redes neurais é o fato de que tal técnica possui capacidade de generalização, ou seja, capacidade de produzir respostas corretas (ou aproximadas) de posse de uma base parcial de conhecimento. Por fim, como a estimativa de qualidade de vídeo é um problema muito abrangente, foram utilizadas múltiplas redes neurais, buscando especializar cada rede neural para um determinado sub-cenário (GoP 10, 20 e 30). Porém para não restringir a proposta para alguns tamanhos de GoPs, uma soma ponderada foi aplicada ao cálculo final da estimativa de qualidade.

O objetivo deste trabalho foi propor um mecanismo denominado MultiQoE que fosse capaz de estimar QoE em tempo real e que possuísse alta correlação com as notas subjetivas (MOS) de vídeos com diferentes níveis de atividade espaço-temporal. Para isto, um conjunto de parâmetros objetivos foram definidos como entrada do sistema. Tais parâmetros foram selecionados após uma análise que comprovasse a influência direta ou indireta na qualidade do vídeo percebido pelo usuário. Dentre os fatores selecionados, alguns possuem grande relação com a qualidade de transmissão (taxa de perda de *frames*) e outros estão mais ligados ao processo de codificação (tamanho do GoP e nível de atividade espaço-temporal). Considerar tais parâmetros diferenciou o MultiQoE dos demais trabalhos relacionados e permitiu ao mecanismo proposto possuir a melhor precisão de estimativa de QoE quando foi utilizada uma base de vídeos com diferentes características de conteúdo e codificação.

Dessa forma, este trabalho apresentou um mecanismo capaz de estimar a qualidade das aplicações de vídeo em tempo real com uma precisão de 90% comparada com a qualidade percebida pelo usuário (MOS). Este desempenho revela a melhor precisão do MultiQoE comparado ao método híbrido PSQA, o qual produziu apenas 41% de acerto. O desempenho do MultiQoE é ainda mais satisfatório quando comparado com as métricas objetivas, visto que tais métricas apresentaram um desempenho de 15, 18 e 27 por cento. Portanto, baseado em tais resultados, é possível concluir que MultiQoE é uma solução aplicável para estimar o nível da qualidade de vídeos em tempo real, assim como, pode ser utilizado em conjunto com mecanismo de otimização e gerência para melhorar o uso dos recursos de rede e o desempenho do sistema.

7.2 **Trabalhos futuros**

Abaixo estão listados os trabalhos futuros a serem realizados:

- Implementar e avaliar MultiQoE em uma rede real e realizar experimentos em maior escala;
- Utilizar outros métodos de avaliação subjetiva, tais como DS;

- Adaptar MultiQoE para aplicações como *games* e vídeos 3D;
- Realizar um estudo comparando o desempenho das múltiplas redes neurais com outras técnicas de inteligência computacional.

7.3 Publicações

Durante o desenvolvimento da proposta desta dissertação os seguintes trabalhos foram submetidos e/ou publicados:

- Publicado: Real-time QoE Prediction for Multimedia Applications in Wireless Mesh Networks (IEEE 4th Future Multimedia Networking (IEEE FMN - IEEE CCNC), 2012)
- Publicado: ETXMULT: a routing metric for multimedia applications in wireless mesh networks (IEEE 3rd Latin American Conference on Communications (IEEE LATINCOM), 2011);
- Publicado: A Hybrid Prediction and Assessment Quality of Experience Approach for Videostreaming Applications over Wireless Mesh Networks (XXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), 2011)
- Submetido: Hybrid Quality of Experience Approach for Real-Time Multimedia Assessment over Wireless Mesh Networks (IEEE Global Communications Conference (GLOBECOM), 2012)
- Submetido: A Real-time Quality of Experience (QoE) estimator for emerging wireless multimedia systems (IEEE Communications Magazine, 2012)

Referências

- [Abdelbaki 1999] Abdelbaki, H. (1999). Random neural network simulator (rnnsim) v. 2. <ftp://ftp.mathworks.com/pub/contrib/v5/nnet/rnnsimv2>.
- [Akyildiz and Wang 2009] Akyildiz, I. and Wang, X. (2009). *Wireless mesh networks*, volume 1. John Wiley & Sons Inc.
- [Basso et al. 1996] Basso, A., Dalgic, I., Tobagi, F., van den Branden Lambrecht, C., Van Den Br, C., and Lambrecht, E. (1996). Study of mpeg-2 coding performance based on a perceptual quality metric. In *In Proceedings of PCS'96*. IEEE.
- [Bouch and Sasse 2001] Bouch, A. and Sasse, M. (2001). Why value is everything: A user-centered approach to internet quality of service and pricing. *Quality of Service—IWQoS 2001*, pages 59–72.
- [Cancela et al. 2008] Cancela, H., Robledo Amoza, F., Rodríguez-Bocca, P., Rubino, G., and Sabiguero, A. (2008). A robust p2p streaming architecture and its application to a high quality live-video service. *Electronic Notes in Discrete Mathematics*, 30:219–224.
- [Cancela et al. 2007] Cancela, H., Rodriguez, P., and Rubino, G. (2007). Perceptual quality in p2p multi-source video streaming policies. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 2780 –2785.
- [Cherif et al. 2011] Cherif, W., Ksentini, A., Negru, D., and Sidibe, M. (2011). Apsqa: Efficient real-time video streaming qoe tool in a future media internet context. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1 –6.
- [CVNI 2010] CVNI (2010). Forecast and methodology, 2009-2014. *White paper, CISCO*, June, 2.
- [da Silva et al. 2008] da Silva, A., Rodriguez-Bocca, P., and Rubino, G. (2008). Optimal quality-of-experience design for a p2p multi-source video streaming. In *Communications, 2008. ICC '08. IEEE International Conference on*, pages 22 –26.
- [Demuth and Beale 1993] Demuth, H. and Beale, M. (1993). Neural network toolbox for use with matlab.

- [Ghareeb et al. 2011] Ghareeb, M., Ksentini, A., and Viho, C. (2011). An adaptive qoe-based multipath video streaming algorithm for scalable video coding (svc). In *Computers and Communications (ISCC), 2011 IEEE Symposium on*, pages 824–829.
- [Ghareeb and Viho 2010a] Ghareeb, M. and Viho, C. (2010a). Hybrid qoe assessment is well-suited for multiple description coding video streaming in overlay networks. In *Communication Networks and Services Research Conference (CNSR), 2010 Eighth Annual*, pages 327–333.
- [Ghareeb and Viho 2010b] Ghareeb, M. and Viho, C. (2010b). A multiple description coding approach for overlay multipath video streaming based on qoe evaluations. In *Multimedia Information Networking and Security (MINES), 2010 International Conference on*, pages 39–43.
- [Ghareeb et al. 2009] Ghareeb, M., Viho, C., and Ksentini, A. (2009). An adaptive mechanism for multipath video streaming over video distribution network (vdn). In *Advances in Multimedia, 2009. MMEDIA '09. First International Conference on*, pages 6–11.
- [Greengrass et al. 2009] Greengrass, J., Evans, J., and Begen, A. (2009). Not all packets are equal, part 2: The impact of network packet loss on video quality. *Internet Computing, IEEE*, 13(2):74–82.
- [Gualdi et al. 2008] Gualdi, G., Prati, A., and Cucchiara, R. (2008). Video streaming for mobile video surveillance. *Multimedia, IEEE Transactions on*, 10(6):1142–1154.
- [Gunawan and Ghanbari 2008] Gunawan, I. and Ghanbari, M. (2008). Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(1):71–83.
- [Hanzo et al. 2007] Hanzo, L., Cherriman, P., and Streit, J. (2007). *Video compression and communications*. IEEE Press.
- [Hashem 1997] Hashem, S. (1997). Optimal linear combinations of neural networks. *Neural networks*, 10(4):599–614.
- [ITU-TP.10/G/100 2007] ITU-TP.10/G/100 (2007). *Appendix I to P.10/G.100: Definition of QoE*. International Telecommunication Union, Geneva, Switzerland.
- [ITUR-B500-11 2002] ITUR-B500-11 (2002). *Methodology for the Subjective Assessment of the Quality of Television Pictures*. International Telecommunication Union, Geneva, Switzerland.
- [Khan et al. 2009a] Khan, A., Sun, L., and Ifeachor, E. (2009a). Impact of video content on video quality for video over wireless networks. In *Autonomic and Autonomous Systems, 2009. ICAS '09. Fifth International Conference on*, pages 277–282. IEEE.
- [Khan et al. 2009b] Khan, A., Sun, L., Jammeh, E., and Ifeachor, E. (2009b). Content classification-based and qoe-driven video send bitrate adaptation scheme. In *Proceedings of the 5th International Mobile Multimedia Communications Conference*, page 16. ICST.

- [Khan et al. 2010] Khan, A., Sun, L., Jammeh, E., and Ifeakor, E. (2010). Quality of experience-driven adaptation scheme for video applications over wireless networks. *Communications, IET*, 4(11):1337–1347.
- [Khashei et al. 2012] Khashei, M., Zeinal Hamadani, A., and Bijari, M. (2012). A novel hybrid classification model of artificial neural networks and multiple linear regression models. *Expert Systems with Applications*, 39(3):2606–2620.
- [Klaue et al. 2003] Klaue, J., Rathke, B., and Wolisz, A. (2003). Evalvid—a framework for video transmission and quality evaluation. *Computer Performance Evaluation. Modeling Techniques and Tools*, pages 255–272.
- [Koumaras et al. 2010] Koumaras, H., Lin, C. H., Shieh, C. K., and Kourtis, A. (2010). A framework for end-to-end video quality prediction of mpeg video. *Journal of Visual Communication and Image Representation*, 21(2):139–154.
- [Lotfallah et al. 2006] Lotfallah, O. A., Reisslein, M., and Panchanathan, S. (2006). A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks. *EURASIP Journal on Applied Signal Processing*.
- [Martinez et al. 2008] Martinez, M., Moron, A., Robledo, F., Rodriguez, P., Cancela, H., and Rubino, G. (2008). A grasp algorithm using rnn for solving dynamics in a p2p live video streaming network. In *Hybrid Intelligent Systems, 2008. HIS '08. Eighth International Conference on*, pages 447–452.
- [Mohamed and Rubino 2002] Mohamed, S. and Rubino, G. (2002). A study of real-time packet video quality using random neural networks. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(12):1071–1083.
- [Moorthy and Bovik 2011] Moorthy, A. and Bovik, A. (2011). Visual quality assessment algorithms: what does the future hold? *Multimedia Tools and Applications*, pages 1–22.
- [Mu et al. 2010] Mu, M., Mauthe, A., and Garcia, F. (2010). A discrete perceptual impact evaluation quality assessment framework for iptv services. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 1505–1510.
- [Mu et al. 2011] Mu, M., Mauthe, A., Haley, R., and Garcia, F. (2011). Discrete quality assessment in iptv content distribution networks. *Signal Processing: Image Communication*.
- [NS2 2012] NS2 (2012). Network simulator 2.34. <http://www.isi.edu/nsnam/ns/>.
- [P.900 2002] P.900, I.-T. (2002). *Subjective video quality assessment methods for multimedia applications*. International Telecommunication Union, Geneva, Switzerland.
- [Par et al. 2006] Par, Y., Seok, Y., Choi, N., Choi, Y., and Bonnin, J.-M. (2006). Rate-adaptive multimedia multicasting over ieee 802.11 wireless lans. In *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*, volume 1, pages 178–182.

- [Piamrat et al. 2009a] Piamrat, K., Ksentini, A., Bonnin, J.-M., and Viho, C. (2009a). Rate adaptation mechanism for multimedia multicasting in wireless networks. In *Broadband Communications, Networks, and Systems, 2009. BROADNETS 2009. Sixth International Conference on*, pages 1–7.
- [Piamrat et al. 2008a] Piamrat, K., Ksentini, A., Viho, C., and Bonnin, J.-M. (2008a). Qoe-aware admission control for multimedia applications in iee 802.11 wireless networks. In *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, pages 1–5.
- [Piamrat et al. 2008b] Piamrat, K., Ksentini, A., Viho, C., and Bonnin, J.-M. (2008b). Qoe-based network selection for multimedia users in iee 802.11 wireless networks. In *Local Computer Networks, 2008. LCN 2008. 33rd IEEE Conference on*, pages 388–394.
- [Piamrat et al. 2009b] Piamrat, K., Viho, C., Bonnin, J.-M., and Ksentini, A. (2009b). Quality of experience measurements for video streaming over wireless networks. In *Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on*, pages 1184–1189.
- [Pinson and Wolf 2004] Pinson, M. H. and Wolf, S. (2004). A new standardized method for objectively measuring video quality. *Broadcasting, IEEE Transactions on*, 50(3):312–322.
- [Rodríguez-Bocca et al. 2007] Rodríguez-Bocca, P., Cancela, H., and Rubino, G. (2007). Video quality assurance in multi-source streaming techniques. In *Proceedings of the 4th international IFIP/ACM Latin American conference on Networking, LANC '07*, pages 83–93, New York, NY, USA. ACM.
- [Seeling et al. 2007] Seeling, P., Fitzek, F. H. P., and Reisslein, M. (2007). *Video traces for network performance evaluation: a comprehensive overview and guide on video traces and their utilization in networking research*. Springer.
- [Seeling and Reisslein 2005] Seeling, P. and Reisslein, M. (2005). Evaluating multimedia networking mechanisms using video traces. *Potentials, IEEE*, 24(4):21–25.
- [Singh et al. 2011] Singh, K., Ksentini, A., and Marienval, B. (2011). Quality of experience measurement tool for svc video coding. In *Communications (ICC), 2011 IEEE International Conference on*, pages 1–5.
- [Vatolin et al. 2009] Vatolin, D., Moskvina, A., Petrov, O., and Trunichkin, N. (2009). Msu video quality measurement tool.
- [Venkataraman and Chatterjee 2011] Venkataraman, M. and Chatterjee, M. (2011). Inferring video qoe in real time. *Network, IEEE*, 25(1):4–13.
- [VTL] VTL. Video trace library. <http://trace.eas.asu.edu/>.
- [Wang et al. 2004] Wang, Z., Bovik, A. C., and Sheikh, H. R. and Simoncelli, E. P. (2004). Image quality assessment: From error measurement to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612.
- [Watkinson 2004] Watkinson, J. (2004). *The MPEG handbook: MPEG-1, MPEG-2, MPEG-4*. Focal Press.

-
- [Yamada et al. 2007] Yamada, T., Miyamoto, Y., and Serizawa, M. (2007). No-reference video quality estimation based on error-concealment effectiveness. In *Packet Video 2007*, pages 288–293. IEEE.
- [Zhang and Hassanein 2010] Zhang, X. and Hassanein, H. (2010). Video on-demand streaming on the internet—a survey. In *Communications (QBSC), 2010 25th Biennial Symposium on*, pages 88–91. IEEE.