



UNIVERSIDADE FEDERAL DO PARÁ

CARLOS JEAN FERREIRA DE QUADROS

CONTROLE DE QUALIDADE PARA 3D-VÍDEO  
*STREAMING* EM REDES EM MALHA SEM FIO

BELÉM, PARÁ

2013

**CARLOS JEAN FERREIRA DE QUADROS**

**CONTROLE DE QUALIDADE PARA 3D-VÍDEO  
*STREAMING* EM 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 Engenharia Elétrica.

Orientador: Eduardo Cerqueira

**BELÉM, PARÁ**

**2013**

Quadros, Carlos Jean Ferreira de .

Controle de Qualidade para 3D-Vídeo *Streaming* em Redes em Malha Sem Fio / Carlos Jean Ferreira de Quadros. – Belém, 2013.

68p.;il.

Orientador: Prof. Dr. Eduardo Cerqueira

Monografia (Computação Aplicada - Redes de computadores e Sistemas Distribuídos) - Universidade Federal do Pará, Instituto de Tecnologia – Programa de Pós-Graduação em Engenharia Elétrica.

1. WMN 2. Qualidade de Experiência 3. Multimídia, FEC. I. Universidade Federal do Pará, Instituto de Tecnologia – Programa de Pós-Graduação em Engenharia Elétrica.

**CARLOS JEAN FERREIRA DE QUADROS**

**CONTROLE DE QUALIDADE PARA 3D-VÍDEO  
*STREAMING* EM 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 Engenharia Elétrica.

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

**BANCA EXAMINADORA**

---

Prof. Dr. Eduardo Cerqueira  
Universidade Federal do Pará - UFPA  
Orientador

---

Prof. Dr. Ádamo Santana  
Universidade Federal do Pará - UFPA

---

Prof. Dr. Agostinho Castro  
Universidade Federal do Pará - UFPA

---

# AGRADECIMENTOS

Em primeiro lugar a Deus por cada dia de vida, a oportunidade de fazer as coisas melhores e da descoberta de novos caminhos e conhecimentos.

À minha amada mãe, Maria Lúcia Ferreira de Quadros e ao meu grande pai Braz Elias da Silva Melo, pelo fundamental apoio que sempre têm me dado nunca desistindo de acreditar em mim.

Aos meus irmãos, Braz Júnior, a minha irmã Fabiane Melo , onde mesmo na distância se fizeram presentes no meu coração.

Ao meu orientador Prof. Dr. Eduardo Cerqueira, que pelo voto de confiança depositado, deu-me a oportunidade de realizar este trabalho. Agradeço pela oportunidade, pela disponibilização dos recursos necessários, o apoio e a ajuda que forneceu para o desenvolvimento e conclusão deste estudo.

Aos amigos queridos, que sempre torceram e acreditaram na realização deste sonho.

Aos professores que contribuíram com minha formação, não só com instrução técnica, mas com princípios que substancialmente se fizeram e fazem importantes para a minha formação.

Aos amigos do GERCOM (Grupo de Estudos em Redes de Computadores e Comunicação Multimídia), pela colaboração significativa com o meu amadurecimento acadêmico e onde tive a oportunidade do desenvolvimento deste trabalho.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior ) pela iniciação científica e o suporte financeiro.

E por fim, mas não menos importante, agradeço também à UFPA, pelo mestrado.

*“O coração é como o relógio da vida. Quem não o consulta, anda naturalmente fora do tempo. Ninguém pode voltar atrás e fazer um novo começo, mas, qualquer um pode começar agora e fazer um novo fim.”*

***Chico Xavier***

---

# SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Objetivos	4
1.2	Contribuições	5
1.3	Estrutura do trabalho	5
<b>2</b>	<b>Tecnologias Utilizadas</b>	<b>7</b>
2.1	Redes em Malha sem Fio	7
2.1.1	Vantagens e Desvantagens	9
2.1.2	Canais sem Fio	11
2.1.3	Métricas e Protocolos de Roteamento	11
2.2	Compressão de Vídeo	13
2.2.1	Princípios de Compressão de Vídeo	15
2.2.2	Estrutura do MPEG-4	17
2.3	Compressão de vídeo 3D	18
2.3.1	Tipos de Representação	18
2.3.1.1	Codificação Convencional de Vídeos Estéreo	19
2.3.1.2	Codificação <i>Video-Plus-Depth</i>	20
2.3.1.3	Codificação de Vídeo <i>Multiview</i> (MVC - <i>Multiview Video Coding</i> )	22
2.4	Conclusões	23

<b>3</b>	<b>Qualidade de Experiência</b>	<b>25</b>
3.1	Introdução à Qualidade de Experiência	25
3.2	Métodos e métricas de estimativa de QoE	26
3.2.1	Métodos Subjetivos	27
3.2.1.1	<i>Single Stimulus (SS)</i>	28
3.2.1.2	<i>Double Stimulus Impairment Scale (DSIS)</i>	29
3.2.2	Métricas Objetivas	29
3.2.2.1	Peak Signal to Noise Ratio (PSNR)	30
3.2.2.2	Structural Similarity Index Metric (SSIM)	31
3.2.2.3	Video Quality Metric (VQM)	32
3.2.3	Métodos Híbridos	32
3.3	Conclusões	32
<b>4</b>	<b>Trabalhos relacionados</b>	<b>34</b>
4.1	Avaliação dos parâmetros	34
4.2	Propostas Relacionadas a Vídeo	35
4.3	Conclusões	37
<b>5</b>	<b>Mecanismo FEC Adaptativo para Vídeo 3D (FECAV3D)</b>	<b>38</b>
5.1	objetivos	38
5.2	Visão geral do Mecanismo FECAV3D	39
5.3	Conclusões	46
<b>6</b>	<b>Simulações e Avaliações</b>	<b>48</b>
6.1	Conclusões	57
<b>7</b>	<b>Conclusões, Trabalhos Futuros</b>	<b>59</b>
7.1	Publicações	61
	<b>Referências</b>	<b>62</b>

---

## LISTA DE FIGURAS

Figura 1	Rede em Malha sem Fio (adaptada de [Wilibox 2009])	8
Figura 2	Representação de um Sistema de <i>Broadcast</i> 3D	15
Figura 3	Diferentes níveis de movimentação de um vídeo	16
Figura 4	Estrutura hierárquica dos quadros do MPEG para vídeos 2D	17
Figura 5	Par de imagens estéreo 3D para cada olho humano e composição da Imagem 3D	18
Figura 6	Predição em H.264/MPEG-4 para Video Convencional Estéreo (retirada de [Smolic et al. 2007])	20
Figura 7	<i>View</i> colorida e o mapa de profundidade respectivo (adaptada de [Song et al. 2012])	21
Figura 8	Estrutura de codificação em camadas para representação <i>Video-Plus-Depth</i>	22
Figura 9	Estrutura de predição temporal/ <i>inter-view</i> para MVC (Retirada de [Smolic et al. 2007])	22
Figura 10	Diferentes escalas da metodologia SS.	28

Figura 11	FECAV3D - Visão geral	40
Figura 12	Árvore das distâncias euclidianas	42
Figura 13	Breakdancers - Tamanho dos quadros x QoE (SSIM)	43
Figura 14	Bullinger - Tamanho dos quadros x QoE (SSIM)	44
Figura 15	Parâmetros de Correção Antecipada de Erros (FEC)	44
Figura 16	Pseudocódigo para o FECAV3D	45
Figura 17	Modelo de Markov com 2 estados	49
Figura 18	Média do MOS por Vídeo 3D	51
Figura 19	Valores médios de SSIM para todas as sequências de vídeo 3D	52
Figura 20	Valores médios de VQM para todas as sequências de vídeo 3D	52
Figura 21	Média do SSIM por Vídeo 3D	53
Figura 22	Média do VQM por Vídeo 3D	54
Figura 23	Sobrecarga de rede	55
Figura 24	Sobrecarga de rede em pacotes	56
Figura 25	Frame 98 original com <i>views</i> esquerda e direita do vídeo Futebol	56
Figura 26	Frame 98 com <i>views</i> esquerda e direita do vídeo Futebol sem FEC	57
Figura 27	Frame 98 com <i>views</i> esquerda e direita do vídeo Futebol com FECAV3D	57

Figura 28 Frame 98 com *views* esquerda e direita do vídeo Futebol com FEC Padrão 57

---

## LISTA DE TABELAS

Tabela 1	Mapeamento do PSNR para o MOS .....	30
Tabela 2	Descrição dos vídeos originais .....	41
Tabela 3	Tabela de notações .....	46
Tabela 4	Valores de ganho em QoE .....	53
Tabela 5	Taxa de recuperação para o vídeo 3D Futebol .....	56

---

## LISTA DE SIGLAS

ACR	Absolute Category Rating
ACEFEC	Adaptive Cross-Layer FEC
AVC	Advanced Video Coding
BCV	Base de Classificação de Vídeo
BFF	Bloco Flexível de FEC
CATT	Contention-Aware Transmission Time
CBR	Constant Bit Rate
CODEC	Coder-Decoder
DCR	Degradation Category Rating
DCT	Discrete Cosine Transform
DPI	Deep Packet Inspection
DSIS	Double Stimulus Impairment Scale
EKS	Encryption Key Sequence
ESF	Extended Subheader Field
ETX	Expected Transmission Count
ETT	Expected Transmission Time
FEC	Forward Error Correction
FEC3D	FEC Adaptativo para Vídeo 3D
FR	Full Reference
FTP	File Transfer Protocol
GoP	Group of Pictures
HCF	Hybrid Coordination Function
HCS	Header Check Sequence
HT	Header Type
HVS	Human Visual System
IEC	International Electrotechnical Commission

IEEE	Institute of Electrical and Electronic Engineers
IP	Internet Protocol
IPTV	Internet Protocol Television
ISO	International Organization for Standardization
ISP	Internet service provider
ITS	Institute for Telecommunication Science
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector
JVT	Joint Video Team
LAETT	Load Aware ETT
MAC	Medium Access Control
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MSE	Mean Squared Error
MVC	Multiview Video Coding
NR	No Reference
NS2	Network Simulator 2
OLSR	Optimized Link State Routing Protocol
OFDM	Orthogonal-Frequency Division Multiplexing
PAD	Proportional Average Delay
PDU	Payload Data Unit
PMP	Point – to – Point
PSNR	Peak Signal to Noise Ratio
QoE	Quality of Experience
QoS	Quality of Service
RR	Reduced reference
RS	Reed-Solomon
RTP	Real Time Protocol
RTSP	Real Time Streaming Protocol
SLA	Service Level Agreement
SS	Single Stimulus
SS	Subscriber Station
SSIM	Structural Similarity Index
SVH	Sistema Visual Humano
TCP/IP	Transmission Control Protocol/ Internet Protocol
TEK	Traffic Encryption Key
TS	Transport Stream
UDP	User Datagram Protocol

UEP	Unequal Error Protection
UGS	Unsolicited Grant Service
VBR	Variable Bit Rate
VCEG	Video Coding Experts Group
VQM	Video Quality Metric
VoIP	Voice over IP
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WMN	Wireless Mesh Network

---

# RESUMO

Redes em Malha sem Fio ( do inglês Wireless Mesh Networks - WMNs) são previstas serem uma das mais importantes tecnologias sem fio no que se refere ao fornecimento do acesso de última milha em redes multimídia futuras. Elas vão permitir que milhares de usuários fixos e móveis acessem, produzam e compartilhem conteúdo multimídia de forma onipresente. Neste contexto, vídeo 3D está previsto atrair mais e mais o mercado multimídia com a perspectiva de reforçar as aplicações (vídeos de vigilância, controle de missões críticas, entretenimento, etc). No entanto, o desafio de lidar com a largura de banda flutuante, escassez de recursos e taxas de erros variantes com o tempo destas redes, ilustra a necessidade da transmissão de vídeos 3D mais resistentes a erros. Dessa forma, alternativas como abordagens de Correção Antecipada de Erros (FEC) se tornam necessárias para fornecer a distribuição de aplicações de vídeo para usuários sem fio com garantia de melhor qualidade de serviço (QoS) e Qualidade de Experiência (QoE). Esta dissertação apresenta um mecanismo baseado em FEC com Proteção Desigual de Erros (UEP) para melhorar a transmissão de vídeo 3D em WMNs, aumentando a satisfação do usuário e permitindo uma melhoria do uso dos recursos sem fio. Os benefícios e impactos do mecanismo proposto serão demonstrados usando simulação e a avaliação será realizada através de métricas de QoE objetivas e subjetivas.

Palavras-Chave: WMN, Qualidade de Experiência , Multimídia, FEC.

---

# ABSTRACT

Wireless Mesh Networks (WMNs) are envisaged to be one of the most important wireless technologies to provide last mile access in future wireless multimedia networks. In this context, 3D-video is envisioned to attract more and more the multimedia market with the perspective for enhanced applications (video surveillance, mission critical control, entertainment, etc.). However, the challenge of dealing with the fluctuating bandwidth, scarce resources and time-varying error rate of these networks, illustrates the need for error-resilient 3D-video transmission. In this context, Forward Error Correction (FEC) approaches are required to provide the distribution of video applications for wireless users with Quality of Service (QoS) and Quality of Experience (QoE) assurance. This study proposal puts forward a FEC-based mechanism with Unequal Error Protection (UEP) to enhance 3D-video transmission in WMNs, while increasing user satisfaction and improving the usage of wireless resources. The benefits and impact of the proposed mechanism will be demonstrated by using simulation, the assessment will be conducted with objective and subjective QoE metrics.

Keywords: WMN, Quality of Experience, Multimedia, FEC.

---

---

# CAPÍTULO 1

---

## INTRODUÇÃO

Recentemente, vê-se o surgimento de tecnologias de redes sem fio de múltiplos saltos, juntamente com a rápida proliferação de uma larga variedade de serviços em tempo real. A Internet está experimentando um considerável crescimento de tráfego que em parte é conduzido por todos esses novos serviços. De acordo com a Cisco, o tráfego IP de vídeo representará mais de 90% do tráfego, IP global em 2015 [Forecast 2010]. Esta perspectiva é facilmente explicada pela grande quantidade de novas formas de informação e entretenimento que são criadas e compartilhadas diariamente por milhares de usuários. Isso inclui conteúdo gerado pelo usuário, sites de notícias, comunidades de redes sociais, *gaming*, materiais de *e-learning*, e serviços atuais como 3DTV e *3D-Video on Demand*.

A indústria 3D, em nível de dispositivos está a cada dia lançando novos aparelhos no mercado. Pouco tempo atrás, era comum haver tecnologia 3D em telas de cinema e TV, mas atualmente, é fácil perceber que está bem mais presente também em notebooks e celulares, por exemplo. Isso, invariavelmente, proporcionará com que usuários consigam e precisem consumir esse conteúdo em diferentes redes o que acaba por gerar mais consumo de banda, interferência ou congestionamento.

Com o crescimento do tráfego de vídeo 3D, torna-se importante garantir uma boa qualidade de transmissão, que pode ser dada por vários fatores. Alguns destes estão relacionados com as características do próprio vídeo, tais como: o tipo de CODEC, *bitrate*, formato de vídeo, tamanho do grupo de imagens (GoP - *Group of Pictures*), conteúdo do vídeo e camadas transmitidas e recuperadas [Yuan et al. 2006]. Desse modo, nem todos os pacotes e camadas têm igual impacto na qualidade aferida [Greengrass et al. 2009]. Há uma forte dependência entre o tipo de informação que o pacote carrega e o impacto que o mesmo tem sobre a qualidade do vídeo percebida pelo usuário.

O maior desafio da distribuição de vídeos 3D, está no fato de requerer a transmissão de pelo menos duas *views* da mesma cena (câmeras levemente afastadas), no caso

do 3D estereoscópico (uma para cada olho, de forma a possibilitar o efeito de imersão e profundidade desejado através da captura da mesma cena por ângulos diferentes). Deste modo, surgiram diversas técnicas de codificação para a compressão de sequências 3D, e entre elas a representação por *video-plus-depth* que comprime um vídeo 3D em duas camadas: um *stream* 2D (mapa de cores) e seu mapa de profundidade sincronizado [Fehn et al. 2004]. Esta é a forma mais comumente usada de representação de vídeos 3D, em virtude da economia da quantidade de dados para a posterior transmissão.

Nesse contexto, em virtude da natureza dos canais de transmissão sem fio, ainda existe o risco de perda de pacotes por interferência ou congestionamento que certamente prejudica a visão e satisfação do usuário final neste tipo de conteúdo. É importante lembrar que em vídeos 3D, a exigência por parte dos usuários vai além do que já se espera para um vídeo 2D, não somente se faz necessária a qualidade da imagem, cores e sucessão de quadros, mas também a qualidade da profundidade percebida pelas pessoas ao assistirem um filme 3D.

Partindo do princípio de que, mesmo se houver perda dos pacotes do mapa de profundidade (camada de reforço), ainda é possível utilizar o mapa de cores (camada base) para formar um vídeo 2D, a importância do mapa de cores torna-se maior em comparação com o mapa de profundidade, ou seja, a camada base possui maior importância do que a camada de reforço. Isso sugere novas formas de otimização baseadas na experiência dos usuários. Em caso de perda de pacotes, pode-se aplicar mecanismos de priorização/descarte/correção de regiões e camadas mais importantes do vídeo para a obtenção de um qualidade final mais aceitável.

Diversas aplicações em variadas áreas já utilizam vídeo 3D em diferentes tarefas. Em determinados serviços como na área de entretenimento, sem um nível adequado de qualidade, a satisfação do usuário pode ficar reduzida ao receber o conteúdo. Na área de saúde, erros médicos podem ser gerados (médicos podem deixar de enxergar/dimensionar problemas de pacientes). Na âmbito da segurança, vigias ou seguranças podem deixar de identificar determinado objeto ou pessoa. Além desses detalhes, esse mesmo conteúdo pode ser de diferentes fontes (*multisource*), por exemplo, eventos como copa do mundo ou alguma outra grande celebração, podem demandar a transmissão de uma grande quantidade de conteúdo para muitos usuários ao mesmo tempo, aumentando o congestionamento. Uma das formas de se resolver esse problema de perda é através de mecanismos de Correção Antecipada de Erros (FEC - *Forward Error Correction*) que insere uma taxa de redundância sobre o conteúdo transmitido. Entretanto o FEC padrão, como conhecido hoje, não é adequado por ser um mecanismo que não foi idealizado e projetado para trabalhar com aplicações específicas como vídeo, bem como vídeos 3D, que não podem receber taxas de redundância às cegas.

Para uma ótima transmissão de vídeos 3D usando mecanismos baseados em FEC, é ideal que seja realizada uma redundância adaptativa, inserindo apenas informações sobre regiões mais importantes e necessárias do vídeo. Com isso, seria possível garantir uma melhor satisfação do usuário, melhorando, por exemplo, as situações citadas anteriormente (médico passe a visualizar melhor, vigia ou segurança identifique melhor objetos, usuários

mais satisfeitos), ao mesmo tempo otimizando os recursos da rede de forma que, com uma redundância mais flexível, seja possível obter também, mais aplicações trabalhando ao mesmo tempo.

Dentro dessas condições supracitadas, existe então a necessidade de um esquema de Proteção Desigual de Erros (UEP - *Unequal Error Protection*), a fim de se priorizar informações mais importantes, para permitir uma boa qualidade de distribuição de vídeo 3D ao introduzir uma quantidade reduzida de informação redundante. Como dito anteriormente, o conteúdo do vídeo também desempenha um papel importante durante a transmissão, tanto para vídeos 2D como para vídeos 3D. Estudos têm demonstrado que vídeos com menor movimentação e complexidade têm melhor resistência à perda de pacotes, o que significa que, para pequenas taxas de perda, a degradação do vídeo será quase imperceptível. Por outro lado, vídeos com alta movimentação ou com alto nível de detalhes são mais suscetíveis, portanto, os defeitos serão mais perceptíveis.

Um importante conjunto de métricas é comumente usado para avaliar a qualidade dos vídeos, possibilitando a identificação e classificação de situações em que ocorrem falhas. A Qualidade de Experiência (QoE - *Quality of Experience*), pode ser definida como a forma como os usuários subjetivamente percebem e avaliam a qualidade de um aplicativo ou serviço [Piamrat et al. 2009]. Isto significa que as análises devem ser feitas fim-a-fim e deve refletir o ponto de vista do utilizador. Para este fim, abordagens objetivas e subjetivas são necessárias.

No que se refere à abordagens subjetivas, estas utilizam observadores humanos para mensurar a qualidade de vídeo, já as abordagens objetivas usam métricas que simulam o sistema de visão humano para identificar e medir estas degradações. Sabe-se que experimentos subjetivos fornecem avaliações de QoE mais factuais, no entanto, este método é demorado, dispendioso e difícil de fazer em tempo real. Assim, as abordagens subjetivas e objetivas são complementares e não excludentes.

Uma vez que grande parte dos serviços de vídeo, 2D e agora 3D, são aplicações em tempo real, estes exigem abordagens QoE *on-the-fly*. Além disso, precisam de um fluxo regular e contínuo dos pacotes, condição que pode sofrer alterações devido a um certo número de fatores típicos de um ambiente sem fio. Em tais redes, as condições do canal podem mudar rapidamente ao longo do tempo, em função de ruído, interferência co-canal, desvanecimento multipercurso, bem como, em virtude de movimento de *hosts* [Lindeberg et al. 2011]. Devido a esses vários desafios, uma Rede em Malha sem Fio (WMN - *Wireless Mesh Network*) foi escolhida para a avaliação do esquema proposto. WMNs surgiram como uma opção relevante para a próxima geração de redes sem fio, pois são auto-configuráveis e fáceis de implantar. Todas essas características proporcionam uma eficiente forma para acesso à Internet de banda larga, e também fornecem uma cobertura flexível e confiável para um grande conjunto de aplicações [Zhu 2011] [Akyildiz and Wang 2005]. No entanto, um dos maiores desafios em WMNs é distribuir de forma justa a largura de banda disponível entre os nós que possuem tráfego em tempo real [Liu and Liao 2009]. Além disso, quando a rede aumenta, o número de transmissões simultâneas aumenta também, causando graves problemas de interferência. Portanto, é desejável otimizar o

uso de recursos para evitar o congestionamento e uma taxa de perda de pacotes elevada, especialmente em aplicações que consomem muitos recursos como vídeo *stream* 3D.

Com o objetivo de superar estes desafios de transmissão, proporcionando uma boa qualidade de vídeo percebida e uma baixa sobrecarga de rede, a adoção de abordagens adaptáveis de proteção de dados torna-se necessária. Esquemas baseados em FEC têm sido utilizados com sucesso em ambientes em tempo real [Nafaa et al. 2008]. Estes, permitem alcançar uma transmissão de vídeo mais robusta enviando dados redundantes juntamente com o conjunto original. Assim, se alguma informação original é perdida, os dados podem ser recuperados através da informação redundante [Lee et al. 2011]. No entanto, como já mencionado, os recursos podem ser limitados e mal distribuídos. Como forma de reduzir a quantidade de informações redundantes, um mecanismo FEC adaptável deve fazer uso de esquemas de UEP. Esta abordagem define a quantidade de redundância dinamicamente, de acordo com a importância dos dados protegidos, permitindo a proteção dos dados mais importantes do vídeo 3D.

Este trabalho propõe um mecanismo *cross-layer* FEC Adaptativo para Vídeo 3D (FECAV3D) com esquema de UEP, que visa assegurar melhor QoE e otimização do uso dos recursos da rede. Devido aos fatos mencionados anteriormente, o uso de mecanismos FEC é adequado para proporcionar melhor qualidade de vídeo. Uma das desvantagens é que para isso é necessário mais largura de banda para enviar os dados de informações redundantes. Para superar este problema, FECAV3D é configurado dinamicamente para enviar informações redundantes só para determinados conjuntos de dados, que causam maior impacto se forem perdidos. A solução proposta é avaliada através de experimentos de simulação com sequências de vídeo 3D reais, usando métricas de QoE subjetivas e objetivas amplamente conhecidas na comunidade acadêmica.

## 1.1 Objetivos

O principal objetivo deste trabalho é avaliar o impacto de técnicas baseadas em FEC adaptativo e não adaptativo sobre a transmissão de vídeo 3D em uma rede com recursos escassos e cenário desafiador (Rede em Malha sem Fio) com base na percepção do usuário final. Este objetivo é alcançado através da utilização do simulador discreto NS2 (*Network Simulator 2*) em conjunto com uma ferramenta para controlar a distribuição de vídeos reais (Evalvid) [Klaue et al. 2003] estendida com módulos para transmissão e avaliação de vídeos 3D. Além disso, propor um mecanismo baseado em FEC adaptativo e UEP, além de medir sua performance a partir de métricas de QoE como VQM (*Video Quality Metric*), SSIM (*Structural Similarity Index*) e MOS (*Mean Option Score*).

Através de simulações, o modelo da distribuição de vídeo sobre a percepção do usuário em ambientes sem fio é investigado, e assim, novas abordagens são propostas para manter aplicações multimídia (vídeo 3D) de tempo real com um nível aceitável em tais redes.

Para isto, a tecnologia WMN, rede de acesso, local e metropolitana sem fio defi-

nida pelo padrão IEEE 802.11s, foi escolhida objetivando a investigação em detalhes do mecanismo proposto. Este padrão 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].

## 1.2 Contribuições

As principais contribuições deste trabalho são:

- A introdução de estratégias avançadas para gerenciamento de vídeo 3D, onde, levando em conta as especificidades dos fluxos de vídeo, é possível melhorar a qualidade de experiência dos usuários e otimizar recursos.
- A identificação dos principais aspectos relacionados com o encaminhamento dos pacotes de vídeo, que se faz necessária tendo em vista a inserção da solução proposta com o propósito do provimento da qualidade do vídeo 3D.
- Através da utilização das métricas de QoE, é feita uma análise da experiência do usuário para o mecanismo proposto e do impacto de parâmetros de codificação do vídeo, definindo-se assim os melhores parâmetros para a obtenção dos melhores resultados e a análise de mecanismos mais eficientes.
- Otimização do código WMN/IEEE 802.11s dentro do ambiente de simulação para suportar mecanismos baseados em FEC adaptativo .

## 1.3 Estrutura do trabalho

O restante do documento é dividido como segue:

- Capítulo 2: É apresentado um estudo acerca da tecnologia de Redes em Malha Sem Fio, ressaltando os conceitos básicos e desafios, além de uma explanação sobre codificação de vídeo 3D.
- Capítulo 3: Apresenta noções de Qualidade de Experiência e as principais métricas existentes para avaliação de vídeo.
- Capítulo 4: Apresenta os trabalhos relacionados com estes conceitos e identifica os principais aspectos que levam em consideração o encaminhamento dos pacotes de vídeo.
- Capítulo 5: Detalha a proposta e apresenta o FECAV3D para melhora da transmissão de vídeo 3D através de um algoritmo baseado em FEC com UEP e que leva em consideração a relevância de cada camada e cada quadro.

- Capítulo 6: Apresenta o ambiente de simulação, avaliação do desempenho e as vantagens/desvantagens do mecanismo proposto.
- Capítulo 7: apresenta as considerações finais sobre os temas abordados na proposta do trabalho, bem como os trabalhos futuros e publicações obtidas no decorrer do desenvolvimento do trabalho.

---

---

## CAPÍTULO 2

---

### TECNOLOGIAS UTILIZADAS

Este capítulo apresenta uma visão geral sobre as tecnologias escolhidas para o desenvolvimento do trabalho e tem o objetivo de fornecer conceitos e definições dentre as 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. Dentre essas pode-se destacar:

1. WMNs, que podem proporcionar uma ampla área de cobertura sem fio e altas taxas de transmissão de dados aos usuários finais;
2. Processo de compressão e representação de vídeos 3D que são desenvolvidos para sua posterior transmissão dentro do cenário WMN;

A Seção 2.1 introduz conceitos importantes sobre as Redes em Malha sem Fio. Na subseção 2.1.1 são apresentadas algumas vantagens/desvantagens sobre WMNs. Uma breve explanação sobre canais sem fio e sobre protocolos e métricas de roteamento são vistos nas subseções 2.1.2 e 2.1.3, respectivamente. Nas seções 2.2 e 2.3 é apresentado uma introdução à compressão de vídeo 2D tradicional e vídeo 3D com os princípios para compressão de vídeo, estrutura do MPEG-4 e os tipos de representação para vídeos 3D.

#### 2.1 Redes em Malha sem Fio

As WMNs surgiram da constatação de que as redes sem fio poderiam ser aproveitadas para reduzir o custo da última milha no acesso à Internet. Através da colaboração entre os nós, um enlace com a rede fixa poderia ser compartilhado, permitindo um uso mais eficiente da banda, evitando assim, o custo da passagem de fios até os usuários finais [Breuel 2004]. O principal atrativo das WMNs é o de prover uma rede comunitária de

acesso, com banda larga e infraestrutura sem fio, oferecendo acesso à Internet com baixo custo [de Albuquerque et al. 2004]. Nestas redes, um cliente *mesh* pode ser de qualquer tipo de dispositivo que tenha conectividade sem fio, ou seja, um notebook, um celular ou um tablet. Normalmente, um ou mais roteadores *mesh*, estão ligados à Internet que servem como portas de entrada para os clientes.

WMNs têm sido usadas como uma possível forma de estender o alcance de pontos de acesso (APs - *Access Points*) sem fio, e têm provado ser uma solução viável para distribuir Internet sem fio confiável em grandes áreas [Pal and Nasipuri 2011]. Este tipo de rede é formada por um conjunto de roteadores *mesh*, geralmente estáticos, que utilizam transmissões *multi-hop* criando um *backbone* e permitindo, desse modo, a comunicação entre os clientes *mesh* [Akyildiz and Wang 2008]. Uma WMN não tem estrutura fixa ou caminho definido e imutável para transmitir as informações. Estas redes contêm caminhos dinâmicos que podem ser alterados e descobertos através de protocolos de roteamento em tempo real, tal como mostrado na Fig. 1 . Ao fazer isso, supõem-se que cada roteador é responsável por manter o fluxo de informações entre outros roteadores próximos.

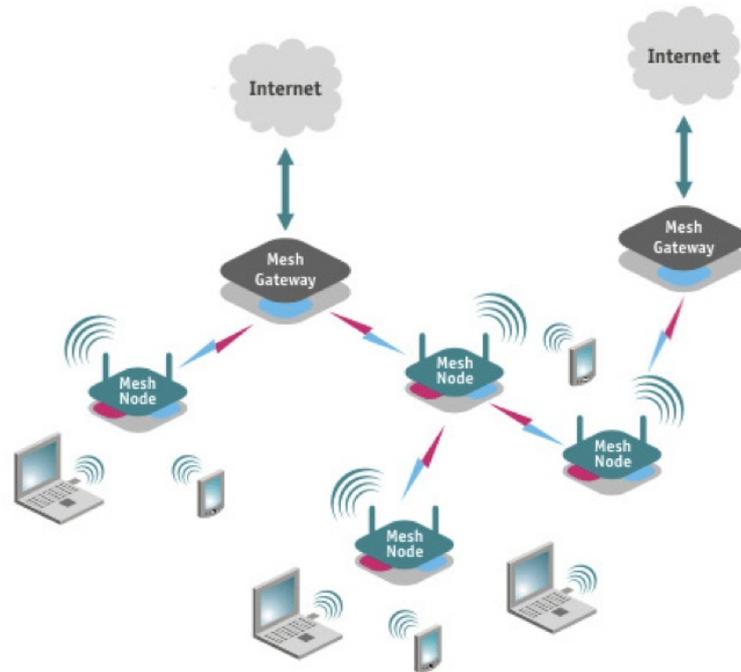


Figura 1: Rede em Malha sem Fio (adaptada de [Wilibox 2009])

Uma grande motivação ao uso das redes em malha é a facilidade e velocidade da sua implantação para ambientes hostis a redes sem fio centralizadas (com muitas barreiras e interferência) e locais onde haja restrição para a instalação de redes com fio (construções históricas e eventos rápidos, por exemplo) [Breuel 2004]. As WMNs também são conhecidas como redes comunitárias de acesso sem fio. Diferentemente das redes infraestruturadas, são redes que dispensam o uso da rede fixa entre os APs. Assemelham-se em muito às redes *ad hoc*, já que ambas utilizam transmissão sem fio e tem topologia dinâmica variável e de crescimento orgânico.

A principal diferença entre as duas tecnologias, no entanto, reside no fato de que nas redes em malha os nós clientes não precisam obrigatoriamente agir como roteadores, possuindo, portanto, menor complexidade nas pontas da rede. Da mesma forma que a rede *ad hoc*, a WMN possibilita a comunicação entre diferentes dispositivos. Alguns participantes irão compor a estrutura principal da rede, trabalhando apenas como roteadores, e comunicando-se via interface sem fio. Outros nós podem se conectar a esses roteadores por cabos e trabalharem apenas como clientes [de Albuquerque et al. 2004].

Para que nós móveis possam fazer parte da rede, geralmente utilizam algum protocolo de roteamento *ad hoc*, para que as rotas sejam alteradas a qualquer momento no caso da mobilidade ou ausência de algum nó. As linhas de pesquisa em WMNs estão diretamente ligadas aos principais desafios da tecnologia, dentre os quais: a busca pelo melhor roteamento na rede, garantir a segurança e integridade da rede dentro do canal sem fio e garantir Qualidade de Serviço (QoS - *Quality of Service*) e QoE perante uma topologia dinâmica e multipercursos. Algumas dessas abordagens podem ser encontradas em [Faccin et al. 2006] e serão discutidas nas subseções 2.1.2 e 2.1.3.

Recentemente, as redes sem fio de múltiplos saltos têm sido consideradas como uma das mais importantes tecnologias sem fio para fornecer acesso de última milha em redes sem fio multimídia futuras. WMNs permitirão que milhares de usuários fixos e móveis passem a acessar, produzir e compartilhar conteúdo multimídia de forma ubíqua. Essa comunicação *multi-hop* é esperada, de forma que atenda aos requisitos de qualidade, a saber (QoS e QoE), garantindo uma fácil implementação, flexibilidade e alta confiabilidade [Bruno et al. 2005] [Zwinkels 2006].

### 2.1.1 Vantagens e Desvantagens

O conceito de WMN traz consigo uma série de vantagens que tornam cada vez mais interessante a sua implantação, como por exemplo:

- Redes de baixo custo: o compartilhamento de recursos faz com que o custo total da rede caia, viabilizando a criação de redes comunitárias [Luiz Jr and Przybysz 2007]. Além disso, pode-se dizer que, como uma WMN pode utilizar também os equipamentos dos próprios usuários como roteadores/repetidores, a necessidade de equipamentos da própria rede diminui sensivelmente. É válido ressaltar que a redução de custo pode ser considerada também pelo fato da ausência de cabeamento;
- Fácil implantação: como as WMNs possuem as características de serem auto-configuráveis, a sua implantação se torna fácil, pois não são necessárias configurações complexas, nem necessidade de mudança caso algum nó venha a entrar na rede;
- Escalável: uma outra característica das WMNs é que sua capacidade de roteamento cresce conforme os nós são adicionados, logo o crescimento das redes, diferente da arquitetura tradicional não se torna um problema [HARADA 2006];

- Otimização do espectro de frequências: considerando que a distância entre os nós diminui sensivelmente, a potência transmitida pode também ser reduzida, permitindo uma maior e mais eficiente reutilização das frequências disponíveis;
- Não há necessidade de linha de visada: com a utilização dos múltiplos saltos entre dois pontos, qualquer exigência de linha de visada entre dois pontos, para uma transmissão de sinais, deixa de existir, pois sempre haverá um caminho que permitirá contornar os obstáculos existentes;
- Tolerância a falhas: a capacidade de roteamento dinâmico, aliado à existência de múltiplas rotas de acesso a um nó, faz com que a rede consiga se recuperar de falhas como a perda de um enlace de comunicação. Não existe um ponto do qual dependa toda a rede. No caso da queda de um nó qualquer, a comunicação passa a ser realizada através de outros nós. Não há a necessidade nem da interrupção de uma comunicação já ativa, pois os próximos pacotes serão roteados através de outros nós alternativos, tornando desta forma, a mudança de rota algo transparente para o usuário;
- Redução da necessidade de conexões entre os APs e a Internet: em uma rede estruturada IEEE 802.11 tradicional, um AP colocado em um *hotspot* apenas para aumentar a capilaridade da rede não necessariamente terá que ter um enlace para a Internet, pois a sua conexão com a rede mundial poderá ser feita através de qualquer nó adjacente.

Entretanto, ainda existem diversas desvantagens, a maioria presente ainda pela recente atenção dada as WMNs por parte tanto do mercado quanto da academia, como por exemplo:

- Interferência: o uso da faixa, não licenciada de 2.4 GHz, possibilita a interferência de equipamentos externos à rede que degradam a qualidade desta como um todo [802.11s 2010];
- Baixo *throughput*: os valores atuais ainda devem ser aprimorados, tendo em mente a possibilidade de crescimento de uma WMN. Uma alternativa para maximizar o *throughput* é utilizar um canal exclusivo para o tráfego de *backbone*, o que gera um desempenho bastante superior [Akyildiz and Wang 2005];
- Falta de segurança: a segurança ainda é um campo aberto no que diz respeito às WMNs, além dos problemas normais de segurança em redes sem fio tradicionais, ainda existe o problema de garantir a privacidade dos dados que estão trafegando entre os nós através de múltiplos saltos [Breuel 2004] [HARADA 2006];
- Ausência de QoS e QoE: assim como os problemas de segurança, a falta de QoS/QoE em WMNs é uma linha de pesquisa que apesar de ser explorada, ainda possui lacunas [Abelém et al. 2007].

## 2.1.2 Canais sem Fio

Considerando nós em uma rede cabeada que transmitem informações usando *links* que têm características estáticas e bem conhecidas, a comunicação entre os nós sem fio ocorre em um ambiente muito mais dinâmico e imprevisível. Um rádio transmite ondas eletromagnéticas através dos canais sem fio em um meio não-homogêneo e desconhecido, em que as propriedades de cada cenário podem ser radicalmente diferentes e oscilam ao longo do tempo [Adlakha et al. 2007]. Estas propriedades podem ser afetadas por objetos próximos, comunicação entre diferentes nós e interferência por radiofrequência em função de outros equipamentos. Isto significa que à medida que o sinal percorre, ele é submetido a uma ampla gama de efeitos físicos que podem provocar uma deterioração da qualidade.

Estas anomalias, por vezes ocorrem como resultado de eventos como desvanecimento por multipercursos e sombreamento que levam a uma diminuição da qualidade do sinal e, assim, tendem a aumentar as taxas de erros de pacotes. Desvanecimento por multipercursos é a consequência de várias reflexões do mesmo sinal. Sombreamento ocorre quando objetos bloqueiam (parcialmente ou não) o caminho de propagação, causando graves atenuações do sinal [Lindeberg et al. 2011].

Além do fato de que um meio compartilhado é usado, as antenas de rádio são incapazes de perceber e transmitir simultaneamente, o que torna as coisas piores. Estas limitações, em conjunto com problemas como terminal oculto (quando um nó é visível a partir de um ponto de acesso, embora não de outros nós conectados ao mesmo AP) e nó exposto (sempre que um nó é impedido de transmitir pacotes devido a transmissões de outros nós vizinhos) aumentam ainda mais as taxas de descarte de pacotes (e perdas) como resultado de um maior número de colisões. Esta situação torna-se ainda pior em redes com alta densidade de nós. Além disso, a medida que a distância entre os nós aumenta, a intensidade do sinal diminui gradualmente.

Assim, devido às características do enlace variáveis com o tempo e da natureza imprevisível dos canais sem fio, são necessários mecanismos para melhorar a transmissão, especialmente devido as exigências rigorosas de *jitter*, atraso e perda de pacote impostas pelos serviços de transmissão de vídeo.

## 2.1.3 Métricas e Protocolos de Roteamento

Devido à natureza dinâmica das WMNs, um protocolo de roteamento é necessário para localizar e manter a conectividade entre os nós da rede, a fim de conseguir o melhor desempenho possível. Em outras palavras, o protocolo de roteamento pode ser descrito como um processo para determinar o percurso fim a fim que liga o nó de origem ao nó-destino [Waharte et al. 2006]. Outro papel deste tipo de protocolo é o de compartilhar informações de roteamento com os nós da rede, como um meio de fornecer o conhecimento necessário para construir uma topologia de rede.

Há uma grande variedade de protocolos de roteamento, mas, geralmente, podem

ser divididos nas categorias: Pró-ativos, Reativos, ou Híbridos. Em protocolos de roteamento Pró-ativos, existe uma troca periódica de mensagens de controle para manter as tabelas de rotas atualizadas mesmo quando não há tráfego de rede, ocorrendo, portanto, uma taxa de desperdício de recursos. Protocolos Reativos ou *On-Demand* foram desenvolvidos para servirem como uma alternativa para as trocas de mensagens constantes. Estes protocolos descobrem rotas em cima dos pedidos, e evitam o desperdício de recursos. O problema que surge com esta abordagem é que existe uma demora inicial maior durante a recuperação de um enlace. Dependendo do tipo de serviço, isto pode não ser aceitável. Por fim, existem os protocolos de roteamento Híbridos, que exploram as vantagens de ambos os protocolos os Pró-ativos e Reativos [Ancillotti et al. 2011].

Os protocolos de roteamento podem usar várias métricas para encontrar a melhor rota fim-a-fim. A contagem de saltos é uma métrica bastante simples, onde apenas o número de saltos entre a origem e o destino são considerados. Uma métrica de desempenho diferente é conhecida como Contador de Transmissão Esperada (ETX - *Expected Transmission Count*) [Couto et al. 2005], que leva em conta a perda de dados e o número de retransmissões. Outra métrica é o Tempo de Transmissão Esperado (ETT - *Expected Transmission Time*) [Draves et al. 2004], que é uma melhoria da métrica ETX, incluindo a largura de banda do enlace para o cálculo do percurso [Parissidis et al. 2011]. Existem outras métricas utilizadas em outros estudos como CATT (*Contention-Aware Transmission Time*) [Genetzakis and Siris 2008], a interferência inter e intra-fluxo [Yang et al. 2006], o número de fluxos admitidos por enlace ( LAETT - *Load Aware ETT* [Aiache et al. 2008]) ou o custo dependente de carga (por exemplo, [Ma and Denko 2007]).

Além destes protocolos e métricas de roteamento, o roteamento baseado em QoS também se torna interessante devido às recentes melhorias em dispositivos móveis, bem como o surgimento de serviços de vídeo em tempo real. Nesta abordagem, as rotas fim-a-fim são determinadas utilizando informações sobre a disponibilidade dos recursos na rede, em conjunto com os requisitos de QoS dos fluxos. Uma característica importante desta abordagem é que possui um mecanismo de detecção de colisão, que lhe permite evitar interferências indesejadas [Bouhouch et al. 2007].

Outra característica importante de alguns protocolos de roteamento para WMNs é o suporte para roteamento *multipath*. Estes algoritmos de encaminhamento aproveitam as várias ligações entre os nós de rede para fornecer mais do que um caminho entre pares de origem-destino [Lindeberg et al. 2011]. Uma vantagem da abordagem de encaminhamento por múltiplos caminhos, é que pode fornecer rotas alternativas de acordo com a QoS desejada [Yang et al. 2006]. Um benefício adicional é a capacidade de transmissão de informações a partir do nó-origem, por meio de uma série de caminhos separados para o nó de destino, utilizando informações de largura de banda, balanceando a carga, reduzindo o atraso e melhorando a tolerância a falhas. A tolerância a falhas é especialmente importante para transmissão em tempo real, já que, quando um bom número de rotas é alcançado, é menos possível que uma ruptura de enlace ocorra em mais de um percurso, ao mesmo tempo, tornando assim, a possibilidade de perda de pacotes menos provável [Tsirigos and Haas 2004] e reduzindo o impacto de uma única falha de enlace sobre a

qualidade de vídeo.

Um protocolo de roteamento eficiente pode melhorar a qualidade de transmissão de vídeo até certo ponto, mas não garante QoE. Por exemplo, se todos os caminhos não cumprirem os requisitos de QoS, se não for rejeitado pelo controle de admissão, a rota menos ruim será escolhida e o vídeo poderá sofrer perda de pacotes, que conduz a deficiências de vídeo. Neste caso, se houver um mecanismo para aumentar a transmissão de vídeo através de códigos de correção de erros, poderá se proporcionar uma melhor QoE percebida pelo usuário final. Em resumo, estas técnicas não são mutuamente exclusivas, e podem proporcionar uma melhoria considerável, se forem devidamente coordenadas.

## 2.2 Compressão de Vídeo

Entre as aplicações da Internet, os serviços que envolvem a transmissão de vídeo 2D e 3D em tempo real têm se tornado cada vez mais populares, como por exemplo, transmissão de eventos esportivos, noticiários, videoconferência, 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 quadros (*frames*) ao usuário. Os quadros 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 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 Electro-*

*technical 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-visual. 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 erros, exijam alto grau de compressão e interatividade como ambientes sem fio.

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-visual foram essenciais para a grande aceitação do MPEG pelo mercado e usuários. Desta forma, assim como para vídeos 2D, a compressão eficiente também é condição essencial para o sucesso de vídeos 3D. Há também, uma forte necessidade para o desenvolvimento de técnicas robustas de compactação de dados nesse tipo de mídia.

Ao longo dos anos, o MPEG-4 sofreu diversas alterações e extensões direcionadas ao tipo de mídia e condições que deve operar. Em 2001 o VCEG (*Video Coding Experts Group*) em parceria com o MPEG desenvolveram o padrão H.264/MPEG-4 Part 10 ou AVC (*Advanced Video Coding*) que é atualmente é um dos formatos mais utilizados para a gravação, distribuição, e compressão de vídeo de alta definição e vídeos escaláveis (*Scalable Video Coding*). O último *draft* sobre a primeira versão da norma foi concluído em Maio de 2003. Em 2006, o JVT (*Joint Video Team*) - resultado da união entre VCEG e MPEG - trabalharam em uma extensão chamada MVC (*Multiview Video Coding*) que inclui as definições e normas para a codificação de vídeos 3D e *free viewpoint television*. O H.264/MPEG-4 Parte 10 é o CODEC utilizado nesse trabalho para a codificação e decodificação das sequências de vídeo 3D, portanto, as seções a seguir darão uma visão geral das técnicas de codificação baseada em *multiview* e profundidade, tendo em vista os disponíveis e emergentes padrões MPEG.

Como apresentado na Fig. 2, 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 (Decoder) atua transformando o formato intermediário em vídeo [Gualdi et al. 2008].

O arquivo intermediário criado pelo codificador é um 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 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, inde-

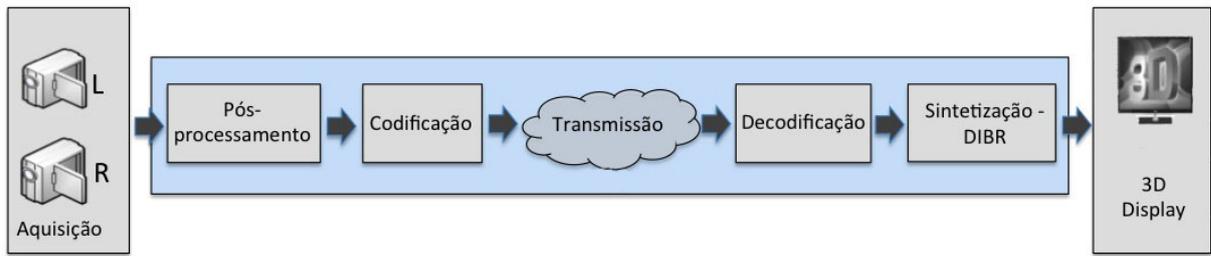


Figura 2: Representação de um Sistema de *Broadcast* 3D

pendentemente 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 utilize seus próprios algoritmos. A única exigência é que a corrente de bits gerada mantenha compatibilidade com o padrão determinado pelo MPEG.

### 2.2.1 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 preditibilidade da informação em questão. Por exemplo, a função seno pode ser predita 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 preditibilidade, 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 predito.

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 2012]. 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 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 é através de canais que sejam 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.

Para vídeos 3D e 2D, o MPEG-4 consegue reduzir as redundâncias espaciais e temporais. A compressão temporal é baseada na técnica de compensação de movimentação. Esta técnica reduz as redundâncias existentes em quadros consecutivos e o principal objetivo é criar um quadros a partir dos quadros anteriores, eliminando as partes em comum dos quadros. Assim, não é necessário enviar cada quadro inteiro, mas apenas as diferenças entre os mesmos.

As Figs. 3(a) e 3(b) exemplificam essa ideia apresentando dois quadros consecutivos de um vídeo. Se a seguinte subtração for feita: quadro 2 - quadro 1, é possível observar o quadro-diferença obtido na Fig. 3(c), onde a porção em preto representa as partes em comum aos dois quadros e as demais partes representam a variação entre os mesmos. Dessa forma, é possível reconstruir o quadro 2 a partir do quadro 1 e do quadro-diferença. De forma similar, a Fig. 3(f) apresenta o quadro obtido a partir das Fig. 3(d) e 3(e). É possível observar que a Fig. 3(f) possui menos partes em preto do que a Fig. 3(c), indicando que este vídeo apresenta um nível maior de movimentação. Além disso, analisando as Figs. 3(f) e 3(c) é possível concluir que a taxa de compressão temporal é menor para vídeos com alta movimentação.

O mesmo raciocínio encontrado na Fig. 3, pode ser utilizado para a representação *multiview* (compressão *inter-view*), conforme será explicado na subseção 2.2.3. A ideia de aproveitar apenas as mudanças entre quadros consecutivos - compressão temporal - pode ser aplicada para aproveitar as mudanças entre duas *views*, ou seja, a imagem do olho esquerdo é codificada usando a imagem do olho direito como referência. Através da compressão *inter-view*, as correlações entre as duas imagens estéreo é explorada, não necessitando a transmissão de duas imagens completas para cada olho.



Figura 3: Diferentes níveis de movimentação de um vídeo

A compressão espacial é definida como sendo técnicas que removem redundâncias de informações presentes nos quadros baseado na propriedade de que os pixels de um quadro estão relacionados com seus vizinhos [Greengrass et al. 2009]. O primeiro passo para a compressão espacial de um quadro é realizar a análise das frequências de tal imagem para encontrar as frequências dominantes. A 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 que 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. O MPEG-4 utiliza a transformada DCT ( *Discrete Cosine Transform* ) por produzir coeficientes apenas da função cosseno e por isso torna-se mais prático [Watkinson 2012].

## 2.2.2 Estrutura do MPEG-4

Para um vídeo 2D, o MPEG define sua estrutura baseada em três tipos de quadros, denominados: I (Intra), P (Preditivo) e B (Bi-direcional). A compressão espacial é aplicada aos macroblocos dos quadros do tipo I, P e B; porém, a compressão temporal é aplicada somente aos macroblocos dos quadros do tipo P e B. Como apresenta a Fig. 4, os macroblocos dos quadros do tipo P utilizam como referência macroblocos dos quadros anteriores do tipo I ou P, já os macroblocos dos quadros do tipo B utilizam como referência macroblocos dos quadros 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 quadros I existirão no vídeo codificado, por exemplo, um vídeo codificado com GoP de comprimento 10 implica que os quadros I irão estar presentes a cada 10 quadros do vídeo. Além do tamanho do GoP existem diferentes padrões internos, os quais são definidos pelo número de quadros P para cada quadro B. Para exemplificar a estrutura de um GoP de vídeo MPEG, a Fig. 4 apresenta uma codificação de vídeo onde o tamanho e o padrão interno do GoP são, respectivamente, 10 e de um quadro P para cada dois quadros B.

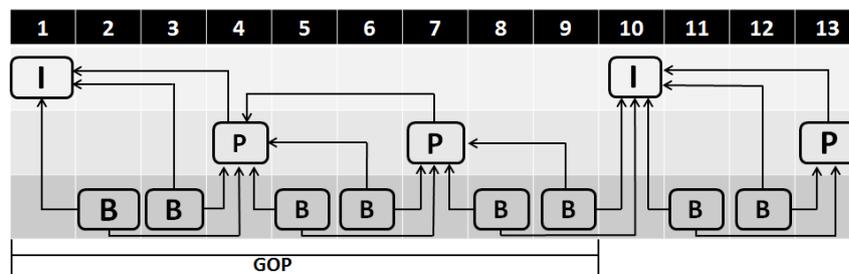


Figura 4: Estrutura hierárquica dos quadros do MPEG para vídeos 2D

A principal consequência do fato do quadro I não realizar compressão temporal é que quando um erro afetar quadros do tipo B ou P, esses erros serão propagados pelos

demais quadros até que um quadro I chegue até o receptor, ou seja, o erro será propagado por todo GoP. No caso da perda em um quadro P, o erro será propagado pelos quadros P e B restantes no GoP. E caso um erro ocorra em um quadro B, este erro não será propagado visto que os quadros B não são usados como referência por outros quadros.

Para um vídeo do tipo 3D existem diversas estruturas que serão apresentadas na seção seguinte. É importante lembrar que as mesmas regras de prioridade, hierarquia e compressão (espacial e temporal) vistas para vídeos tradicionais 2D, são também adequadas para os formatos que serão mostrados a seguir, conforme será explicado.

## 2.3 Compressão de vídeo 3D

Atualmente, a tecnologia 3D para vídeos, em sua grande maioria, emprega a estereoscopia, utilizando algum dispositivo de visualização especial para projetar vídeos em um campo tridimensional. Imagem estereoscópica, ou imagem estereoscópica 3D é uma técnica capaz de gravar e apresentar informação visual tridimensional ou criar a ilusão de profundidade em uma imagem [Kalva et al. 2006]. A ilusão de profundidade numa fotografia, filme, ou em uma imagem bidimensional é criada através da apresentação de uma imagem ligeiramente diferente para cada olho. Muitas telas 3D usam este método para apresentar imagens, conforme mostrado na Fig. 5. No lado do usuário, é necessário um óculos adequado à técnica de reprodução utilizada: anaglifo, dolby 3D, polarizado ou frame-sequencial (que é o caso deste trabalho).



Figura 5: Par de imagens estéreo 3D para cada olho humano e composição da Imagem 3D

### 2.3.1 Tipos de Representação

Muitos sistemas 3D são baseados em cenários, onde a cena 3D é capturada por um certo número de câmeras. O caso mais simples é o vídeo estéreo clássico com duas imagens, uma para cada olho. Sistemas mais avançados aplicam 8, 16 ou mais câmeras. Alguns sistemas são baseados em dados de profundidade que também podem ser tratados como sinais de vídeo. Esta subseção apresenta uma visão geral de algoritmos e padrões de compressão para tais dados. Uma visão inicial desta área de pesquisa pode

ser encontrada em [Shum et al. 2003]. Dependendo do grau de similaridades compartilhado por um subconjunto do câmeras, um ganho de compressão pode ser alcançado em relação a codificação de cada uma das cenas separadamente. Na codificação *multiView*, correlações entre câmeras adjacentes são exploradas, além de correlações temporais dentro de cada seqüência. Portanto, a codificação *multiview*, acrescenta uma outra dimensão de compressão: a compressão *inter-view*.

### 2.3.1.1 Codificação Convencional de Vídeos Estéreo

A codificação de pares de vídeo estereo é o caso mais importante e especial de *multiview* com  $N = 2$  *views*. A compressão de vídeo estereo convencional tem sido estudada há muito tempo, e as normas correspondentes estão disponíveis. Um par estereo convencional consiste de duas imagens que mostram a mesma cena a partir de dois pontos de vista ligeiramente diferentes, correspondentes à distância dos olhos humanos. As imagens são em geral muito semelhantes, o que os torna adequados para a compressão com uma imagem predizendo a outra. Por exemplo, uma delas pode ser comprimida sem referência à outra imagem. Em seguida, a segunda imagem pode ser predita a partir da que já foi codificada, tal como imagens temporalmente relacionadas podem ser comprimidas usando a compensação de movimento do vídeo [Smolic et al. 2007].

A disparidade de cada imagem estereo com relação a outra é equivalente a disparidade de movimentação entre duas imagens consecutivas de uma seqüência de vídeo. Portanto, justifica-se utilizar os mesmos princípios de compensação de movimentação vistos na compressão temporal para a estimativa e compensação de disparidade entre as duas *views* estereo. Em seguida, resta apenas a codificação da diferença residual entre estas.

No entanto, algumas diferenças específicas entre a compensação de movimentação e a compensação de disparidade necessitam ser consideradas. As estatísticas dos vetores de disparidade são diferentes das estatísticas dos vetores de movimentação. Os vetores de disparidades entre *views* estereo podem assumir valores relativamente grandes ou pequenos mas seguindo um padrão. Disparidade Zero significa uma profundidade muito grande do ponto 3D correspondente, enquanto que pontos 3D próximos da câmera podem ter valores muito grandes de disparidades. Isso pode exigir ajustes de codificação de entropia dos vetores de disparidade. Em geral, imagens temporalmente adjacentes de uma seqüência de vídeo tendem a ser mais semelhantes do que os dois pontos de vista de um par estereo. Efeitos de desocclusão, isto é, conteúdos que são visíveis em uma imagem e não visíveis na outra, não podem ser previstos e são, em média, mais evidentes em um par estereo do que entre duas imagens de vídeo adjacentes temporalmente.

A combinação de previsão *inter-view* e temporal é o princípio básico para a compressão de vídeo estereo convencional. A especificação correspondente padrão já foi definida no ITU-T Rec. H.262/ISO/IEC 13818-2MPEG-4. O Perfil *Multiview* [Haskell et al. 1996], é ilustrado na Fig. 6. A visão do olho esquerdo é codificada sem referência à visão do olho direito, usando o padrão MPEG-4. Desse modo, é possível decodificar o fluxo de bits do olho esquerdo, permitindo a visualização e exibição de um vídeo 2D.

Para a visão do olho direito, a predição *inter-view* é realizada juntamente com a predição temporal.

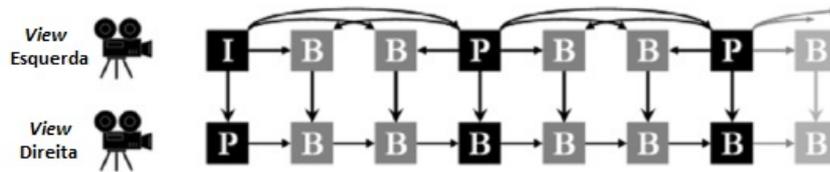


Figura 6: Predição em H.264/MPEG-4 para Video Convencional Estéreo (retirada de [Smolic et al. 2007])

No entanto, o ganho de eficiência em comparação com a codificação de compressão independente de ambos os fluxos de vídeo é bastante limitado. Isto ocorre devido ao fato de que a predição temporal já proporciona um desempenho muito bom. Normalmente, se a predição temporal é eficiente para uma determinada imagem (por exemplo, quadros B para a vista direita na Fig. 6, então a predição *inter-view* não aumenta o desempenho de codificação de forma significativa. Imagens temporalmente vizinhas geralmente são mais semelhantes que imagens de *views* diferentes, como dito antes.

A pesquisa em compressão de vídeo estéreo convencional continua em várias direções, no entanto, nenhum dos desenvolvimentos incluindo o perfil de *Multiview* original têm alcançado relevância comercial até agora, já que, a aplicação de vídeo estéreo não se transformou em um mercado de massa relevante ainda.

### 2.3.1.2 Codificação *Video-Plus-Depth*

Uma alternativa para vídeo estéreo convencional, visto na seção anterior é a de transmitir um sinal de vídeo e um mapa de profundidade sincronizado. A partir das informações do vídeo e profundidade, um par estéreo pode ser construído no decodificador [Fehn et al. 2002] [Fehn 2004]. Isso amplia as funcionalidades, uma vez que permite a visualização do efeito 3D mesmo havendo a movimentação da cabeça do usuário (autoestereoscopia), se o movimento deste é monitorado. Além disso, este formato é interessante do ponto de vista da eficiência de compressão. Por exemplo, dados de profundidade podem ser considerados como um sinal de vídeo monocromático, de luminância somente. A profundidade é dada pelo intervalo entre dois pontos que indicam a distância mínima e máxima da cena 3D correspondente a partir da câmera. Ou seja, o intervalo de profundidade é linearmente quantizado com 8 bits, ou seja, o ponto mais próximo é associado com o valor de 255 e o ponto mais distante, é associado com o valor 0. A informação de profundidade é especificada, resultando em uma imagem em escala de cinza conforme mostra a Fig. 7 .

Os resultados do projeto europeu ATTEST [Fehn et al. 2002], mostraram que os dados de profundidade podem ser comprimidos de forma muito eficiente. Vários codecs do estado-da-arte foram testados (MPEG-2, MPEG-4, H.264/AVC) e uma estimativa indica

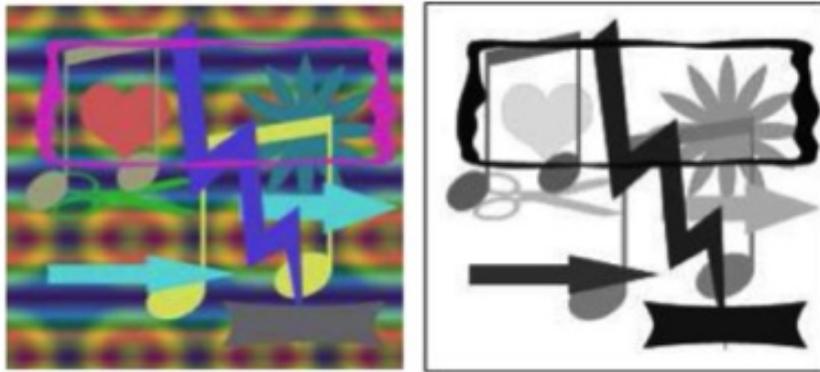


Figura 7: *View* colorida e o mapa de profundidade respectivo (adaptada de [Song et al. 2012])

que 10% a 20% da taxa de bits que é necessário para codificar um vídeo é suficiente para codificar o mapa de profundidade em boa qualidade. Isto acontece devido aos dados específicos de dados de profundidade, serem, em média, mais suaves e menos estruturados que os dados de cor.

Um problema geral da representação *video-plus-depth* é a criação de conteúdo, ou seja, a geração de informações de profundidade. As câmeras que capturam automaticamente pares de profundidade com o vídeo estão disponíveis e estão sendo desenvolvidas, mas a qualidade dos mapas de profundidade capturados ainda é limitada. Algoritmos para estimativa de profundidade têm sido estudados exaustivamente na literatura e novas soluções têm surgido. No entanto, ainda permanecem apenas estimativas que só podem ser resolvidas se houver uma probabilidade de erro residual. Erros de estimativa influenciam a qualidade de *views* renderizadas. Um automático, preciso e confiável sistema de captura de profundidade ainda será desenvolvido. Com esta representação, a exigência de largura de banda é menor do que a codificação independente das *views* (cerca de 1,2 vezes do vídeo 2D), mas isso requer uma renderização adicional da imagem de profundidade no receptor para construir o fluxo de vídeo 3D.

Neste trabalho foi utilizada a abordagem de codificação em camadas (*Layered Coding Approach*) [Chung 2010]. Por razões de clareza, a Fig. 8 ilustra alguns conceitos da representação *video-plus-depth* que utiliza esta abordagem. Pode-se observar que os dois GOPs (superior e inferior) formam um par, o primeiro é a camada de base (*base layer*) e o de baixo é a camada de reforço (*enhancement layer*). A codificação é composta dos mesmos três tipos de quadros vistos na seção 2.2.2, designados por I, P e B. Quadros I são codificados por meio de compressão espacial e sem nenhuma referência a outros quadros da sequência. Para conseguir a compressão temporal, os quadros P são reconstruídos com a ajuda da compensação de movimentação a partir do quadro I ou P anterior. Como resultado, quadros P têm uma relação de compressão melhor do que os quadros I, mas isso também depende da quantidade de movimentação presente na sequência. Quadros B possuem um melhor índice de compressão utilizando a predição do último e do próximo quadro I ou P, conforme visto antes.

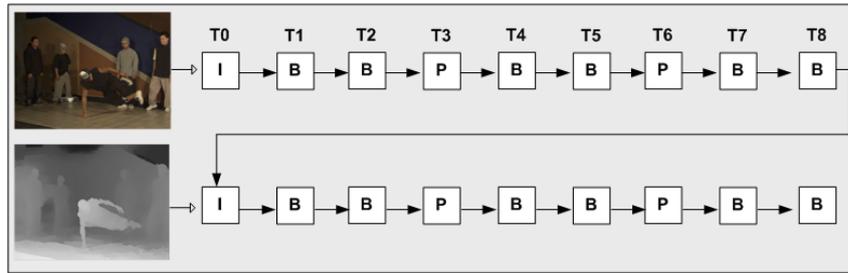


Figura 8: Estrutura de codificação em camadas para representação *Video-Plus-Depth*

### 2.3.1.3 Codificação de Vídeo *Multiview* (MVC - *Multiview Video Coding*)

Um elemento comum de muitos sistemas 3D é o uso de várias *views* da mesma cena que devem ser transmitidas para o utilizador. A solução simples e direta para isso seria codificar todos os sinais de vídeo de forma independente usando algum codec de vídeo do estado da arte, como H.264/AVC. No entanto, vídeo *multiview* contém grandes quantidades de dependências estatísticas *inter-views*, uma vez que todas as câmeras capturaram a mesma cena a partir de diferentes pontos de vista. Estas dependências podem ser exploradas através de predição de redundância temporal/*inter-view*, tal como ilustrado na Fig. 9. As imagens não são apenas previstas a partir das imagens temporalmente vizinhas, mas também a partir de imagens correspondentes nas vistas adjacentes. Avaliações estatísticas mostram que um ganho significativo pode ser esperado da predição temporal combinada com a predição *inter-view* [Merkle et al. 2005] [Kaup and Fecker 2006].

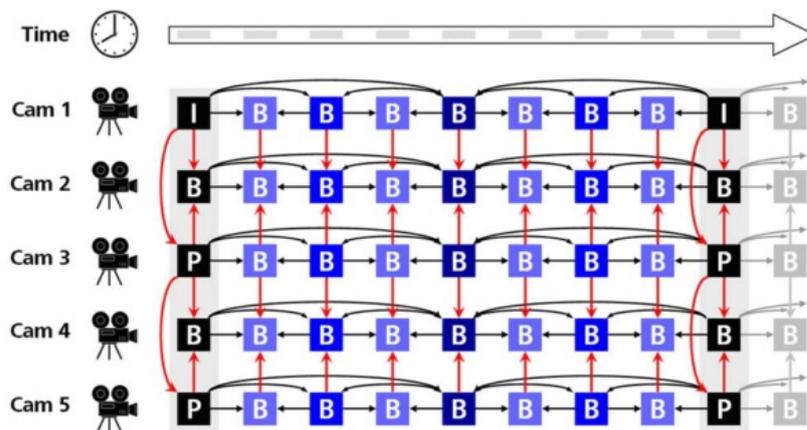


Figura 9: Estrutura de predição temporal/*inter-view* para MVC (Retirada de [Smolic et al. 2007])

Vários grupos tem dirigido sua pesquisa para a codificação de vídeo *Multiview*, dedicando-se a estruturas de predição *inter-view*/temporal para a exploração de formas eficientes das dependências estatísticas dentro dos conjuntos de dados *multiview* [Oh and Ho 2006]. Entre estes, algoritmos baseados em hierarquia de quadros B [Schwarz et al. 2006], conforme suportado pela sintaxe do H.264/AVC na predição temporal e *inter-view* (Fig. 9) revelaram melhor desempenho em experiências exaustivas [Flierl et al.

2007]. Nestas experiências, foi demonstrado por meio de medições objetivas e subjetivas que codificação MVC supera codificação independente dos fluxos de vídeo múltiplos de forma significativa. No entanto, o ganho depende fortemente do conteúdo e das suas propriedades, como a distância da câmera, taxa de quadros e complexidade do conteúdo (movimento, textura). Para alguns conjuntos de dados o ganho do *Peak Signal to Noise Ratio* (PSNR) foi dado como de 0,5 a 3 dB.

Uma desvantagem da combinação de predição temporal/*inter-view*, como ilustrado na Fig.9 é a complexidade. Isto inclui complexidade computacional, requisitos de memória e de atraso. Em [Mueller et al. 2006], foi demonstrado que a complexidade pode ser significativamente reduzida sem sacrificar muito a eficiência da codificação. Predição *inter-view* é limitada para frames que deveriam ser tratados como frames chave (quadros I), o que não ocorre na codificação independente dos pontos de vista. A maior parte do ganho da codificação MVC vem da predição *inter-view* destes quadros que não utilizam o predição temporal por razões de acesso aleatório. Omitir a predição *inter-view* de imagens que tem referência temporal não custa muito da eficiência da codificação e diminui a complexidade de forma significativa.

Pesquisas sobre codificação estéreo de vídeo, vídeo *multiview* e *video-plus-depth* atingiram um bom nível de maturidade. Normas internacionais disponíveis, permitem uma grande variedade de sistemas 3D e aplicações. No entanto, em comparação com outros tipos de codificação de dados de mídia do campo científico, ainda é relativamente jovem, portanto, ainda há muito espaço para melhoria dos algoritmos. Isto inclui, por exemplo, otimização do MVC e desenvolvimento de novos algoritmos de codificação MVC. Codificação de mapas de profundidade ou disparidade podem ser melhoradas ainda mais, pelo desenvolvimento de algoritmos dedicados. Além disso, há tipos mais complexos de representações de dados para vídeos 3D, como vídeo com profundidade em camadas (*layered depth video* e vídeo *multiview* mais profundidade (*multiview video plus depth*) que fornecem funcionalidade estendidas e que ainda devem ser aprimorados, pois, algoritmos eficientes de codificação para esses dados ainda estão sob investigação. Neste trabalho a representação por *video-plus-depth* foi utilizada para a transmissão das sequências de vídeo 3D. Deve-se notar que para quaisquer tipos de codificação, o mecanismo proposto 3DV-FEC pode ser facilmente adaptado em virtude da sua natureza modular, conforme será visto no capítulo 5.

## 2.4 Conclusões

Este capítulo introduziu as principais tecnologias e conceitos usados neste estudo. Levando em consideração, as especificidades de cada uma, foi apresentada uma breve explanação acerca dos principais tópicos que norteiam as funcionalidades de cada tecnologia, necessários para o futuro entendimento do mecanismo proposto no decorrer do trabalho.

A implantação generalizada de WMNs, bem como o uso de dispositivos móveis, tem aumentado significativamente nos últimos anos. Além disso, vários novos serviços

têm exigências rigorosas para trabalhar em tempo real. Esta situação é uma causa de preocupação no que diz respeito à transmissão de vídeo 3D e 2D, que pode ser considerada como sendo um dos serviços mais exigentes em termos de atraso, taxa de pacotes, perda e largura de banda. Ambientes que não oferecem as condições necessárias têm mau desempenho neste tipo de serviço e deficiências de vídeo são muito visíveis para os usuários finais.

Neste capítulo, componentes importantes de transmissão de conteúdo de vídeo, WMNs, canais sem fio, características de protocolos de roteamento e detalhes de codificação de vídeo foram explorados. Além disso, vários fatores que podem causar degradação da qualidade durante a transmissão de vídeo também foram analisados. Após investigar todas as questões relacionadas, a melhoria da qualidade de transmissão de vídeo 3D e 2D continua a ser uma tarefa muito desafiadora por conta do grande número de componentes interferentes e as relações complexas entre eles.

O padrão H.264/MPEG-4 Parte 10 utiliza técnicas de compressão temporal, espacial e *inter-view* durante o processo de codificação do vídeo 3D. Além disso, 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.

Foi estabelecido no desenvolvimento do estudo, a mesclagem da tecnologia de redes em malha sem fio (WMN) com tráfego de vídeo 3D em tempo real, usando a representação *video-plus-depth* para a aferição de melhorias no desempenho do mesmo sob a ótica de métricas de Qualidade de Experiência que serão apresentadas no próximo capítulo.

---

---

## CAPÍTULO 3

---

# QUALIDADE DE EXPERIÊNCIA

Vídeos estereoscópicos estão em crescente demanda e a disponibilidade tem aumentado em função dos avanços em tecnologias de visualização, aquisição, transmissão e eficientes formas de compressão. Portanto, a necessidade de avaliação da qualidade em vídeos 3D acaba se tornando importante para a melhoria da satisfação do usuário e gestão eficiente dos recursos.

Deste modo, o estudo da Qualidade de Experiência (*Quality of Experience* - QoE) tornou-se fundamental para avaliação da qualidade de aplicações de vídeo 2D e 3D e vem ganhando impacto na comunidade de processamento e distribuição dessas mídias. Isto levou ao surgimento de diferentes métricas objetivas para imagens e vídeos. Este capítulo objetiva apresentar as principais definições e abordagens de QoE. A Seção 3.1 apresenta os principais conceitos sobre QoE, já 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 [Rec 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 3D tornam as aplicações de vídeo extremamente sensíveis à perdas de dados, devido à estrutura de interdependência existente entre os quadros de um vídeo 3D comprimido. Dessa forma, estas aplicações 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 3D 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 estimatimam 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 como: 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 3D 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 3D transmitidos em tempo real em redes de computadores, visto que estas 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 RR 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 exigiria muitos recursos (humanos, tempo) durante o processo de realização dos testes subjetivos.

Na comunidade científica há especificações bem definidas para realização das avaliações subjetivas 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 apresentados 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. 10. É possível observar que a Fig. 10.a apresenta a escala com 5 diferentes níveis de qualidade, porém a escala de qualidade presente na Fig. 10.c lista 5 níveis de danos. Por outro lado as Fig. 10.b, Fig. 10.f e Fig. 10.e apresentam escalas contendo 9, 11 e 7 níveis de qualidade, respectivamente. Por fim a Fig. 10.d apresenta uma escala de qualidade contínua onde é possível atribuir uma nota entre 100 e 0.

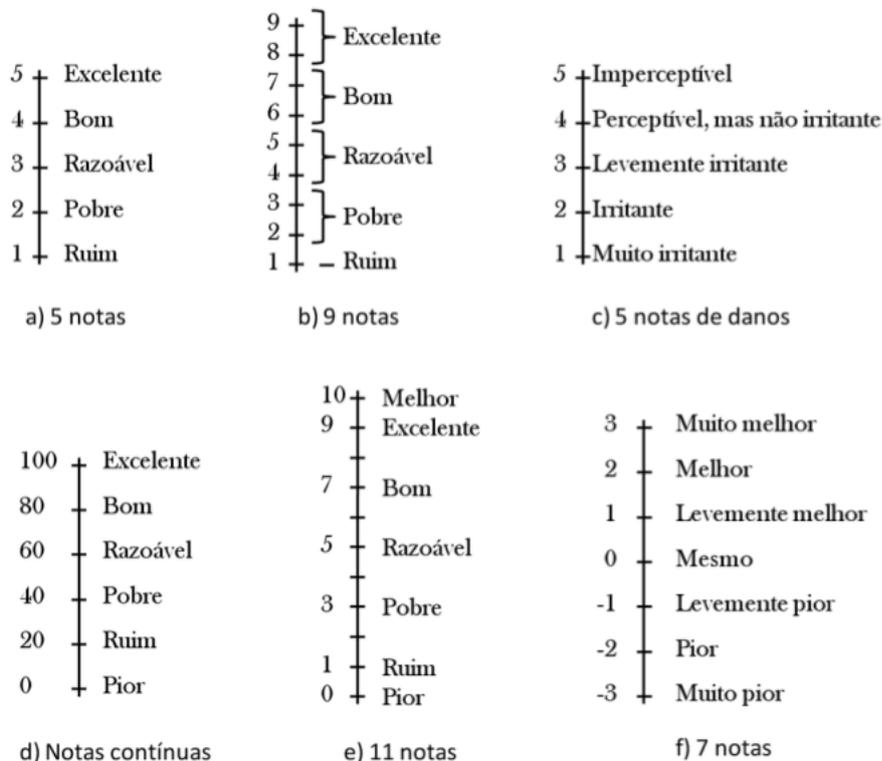


Figura 10: Diferentes escalas da metodologia SS.

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 seja 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 esclarecerem dúvidas sobre a avaliação. 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] e *Structural Similarity Index Metric* (SSIM) [Wang et al. 2004b].

### 3.2.2.1 Peak 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:

$$MSE = \frac{1}{MN} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2 \quad (3.1)$$

Onde:

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

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

$K(i, j)$  : 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 3.2:

$$PSNR = 10 \log_{10} \left( \frac{MAX_1^2}{MSE} \right) \quad (3.2)$$

Onde:

$MAX$  : Valor máximo de luminosidade

É possível observar na Equação 3.2 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. 10.a.

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)

Porém o PSNR desconsidera fatores importantes do sistema visual humano e a crominância dos *pixels*, o que impossibilita o uso desta métrica em vídeos 3D, em virtude da sua inutilização para avaliação dos mapas de profundidade. Além disso, a estimativa de qualidade fornecida por este método é 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.

### 3.2.2.2 Structural Similarity Index Metric (SSIM)

Diferentemente do PSNR que somente compara a taxa de erro do vídeo recebido em relação ao vídeo original, a métrica SSIM avalia o vídeo recebido considerando outros fatores como o SVH (Sistema Visual Humano). Surgindo em função de o SVH ser altamente eficiente em extrair informações visuais das imagens/vídeos e não das taxas de erro, o SSIM analisa similaridade de cores, luminosidade e estrutura [Uemura et al. 2008]. Os valores extraídos do frame recebido pelo usuário e do frame original são armazenados em vetores separadamente, um vetor para luminosidade, outro vetor para estrutura e um vetor para cor. Posteriormente, obtém-se a média de cada vetor e a combinação dessas três médias gera o valor do SSIM, indicando a qualidade do vídeo. O valor de SSIM é obtido através da Equação 3.3 [Gao et al. 2009]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (3.3)$$

Onde:

$\mu_x$  é média de x;

$\mu_y$  é a média de y;

$\sigma_x$  é o desvio padrão de x;

$\sigma_y$  é o desvio padrão de y;

$\sigma_{xy}$  é a covariância de y ;

$c1 = (k_1L)^2, c2 = (k_2L)$  são duas constantes ;

L = O valor máximo que pode ser atribuído a cada pixel;

$k_1 = 0,01$  e  $k_2 = 0,03$  por padrão.

O valor de SSIM é expresso como um valor decimal entre 0 e 1 ( quanto mais próximo de 0, pior é a qualidade do vídeo; quanto mais próximo de 1, melhor é a qualidade do vídeo, no caso em que o valor é igual a 1, o vídeo recebido pelo usuário corresponde em 100% ao vídeo original ) [Takahashi et al. 2008].

### 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. Esta métrica é mais completa que PSNR e SSIM, pois avalia a distorção das cores, dos pixels, ruído e se o vídeo está “borrado”.

A métrica, assim como as demais, baseia-se na comparação do vídeo recebido pelo usuário em relação ao vídeo original, comparando o brilho e o contraste. A primeira etapa consiste na “calibração”. Nessa etapa, o contraste e o brilho dos vídeos são comparados. A segunda etapa consiste na comparação do alinhamento espacial. Verifica-se a posição no eixo horizontal e vertical de ambos os vídeos. A Terceira etapa consiste em analisar a quantidade de pixels perdidos em relação ao vídeo original. A quarta etapa refere-se ao alinhamento temporal que compara o tempo no qual cada frame foi recebido em relação ao tempo no qual o mesmo deveria ter sido recebido. Esta etapa consiste em avaliar o atraso do vídeo recebido pelo usuário.

O VQM é dado por um número real, onde, quanto mais próximo o valor for de 0, melhor será a qualidade do vídeo, o que indica uma menor distorção em relação ao vídeo original [Vercammen et al. 2008].

### 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 originais e degradados 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*.

## 3.3 Conclusões

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.

A investigação em curso tem-se centrado sobre o uso de métricas de avaliação objetivas para vídeos 3D. Métodos objetivos, como o PSNR não podem dar nenhuma informação sobre a percepção de profundidade e, portanto, não podem ser usados para esses tipos de vídeos. Neste trabalho foram utilizadas as métricas objetivas SSIM e VQM, além de testes subjetivos para a confirmação dos resultados do mecanismo proposto.

---

---

# 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 propõem mecanismos que sejam capazes de fornecer um aprimoramento de QoE em tempo real, já que esta tarefa é complexa por existirem muitos fatores que, combinados, afetam a qualidade percebida de um vídeo. Além disso, poucos trabalhos propõem estudos e melhorias para serviço de vídeos 3D *streaming* que são emergentes. Este capítulo apresenta os principais estudos relacionados com o objetivo proposto neste trabalho, incluindo mecanismos de correção de erros e otimização do nível de QoE.

### 4.1 Avaliação dos parâmetros

Diversas abordagens relacionadas com soluções para melhoria de tráfego de vídeo já foram propostas, cada uma com suas funcionalidades e objetivos definidos. Entretanto, alguns requisitos são sugeridos para que estes mecanismos funcionem de forma satisfatória e alcancem um desempenho ótimo:

- Levar em consideração a percepção do usuário final, ou seja, diferentes pacotes devem receber diferentes taxas de prioridade de acordo com a contribuição que oferecem para a aferição de Qualidade de Experiência;
- Otimização dos recursos da rede por parte dos mecanismos de controle de pacotes;
- Ser escalável: independente do número fluxos que atuem na rede, o mecanismo deve funcionar sem perder o desempenho;

- Ser flexível, ou seja, funcionar independentemente da tecnologia, arquitetura de rede ou parâmetros de codificação (valores de tamanho de GOP, *bitrate* ou número de quadros B);
- Manter a interoperabilidade: nós que usam algoritmos de otimização e seleção, devem ainda ser capazes de se comunicar com nós sem estes mecanismos;
- Minimizar o número de mudanças: as alterações em dispositivos de rede e terminais de usuário finais devem ser reduzidas;
- Respeitar a prioridade dos fluxos: fluxos mais importantes para os usuários e com maior prioridade, como os fluxos de voz, não deve ser prejudicados;

Além disso, com a proliferação da comunicação em redes sem fio e das diversificadas tecnologias de acesso, é importante levar em consideração a mobilidade do usuário e o impacto da mesma. Finalmente, as avaliações devem estar de acordo com requisitos e métricas de QoE e do Sistema Visual Humano (HVS - *Human Visual System* ).

## 4.2 Propostas Relacionadas a Vídeo

Diversos mecanismos de controle para a transmissão de fluxos de vídeo e otimização dos níveis de QoE foram e são desenvolvidos para tecnologias com e sem fio. Estudos como Descarte Seletivo de Pacotes, Marcação de Pacotes, Estimativa de QoE, Correção de Erros, entre outros, foram e são propostos. Os trabalhos citados a seguir apresentam recentes avanços e focam na criação de técnicas e métodos para otimização dos fluxos de vídeo e gerenciamento da transmissão.

O mecanismo *Adaptive Cross-Layer FEC* (ACFEC) usa correção de erros em nível de pacote [Han et al. 2010]. Através de uma concepção *cross-layer*, estes pacotes são monitorados na camada MAC, e o número de pacotes FEC redundantes é aumentado ou diminuído dependendo de um contador de falhas. Desta forma, quando o contador é zero, significa que não há perda de pacotes e a conexão é boa, assim, os pacotes de recuperação não são gerados o que resulta em menos tráfego redundante. No entanto, nenhuma avaliação da sobrecarga da rede é conduzida.

Como mencionado anteriormente, redes em malha sem fio não possuem geralmente uma alocação de banda justa, que significa que o aumento da sobrecarga de um nó pode afetar a comunicação entre os outros. Além disso, a abordagem mencionada não considera o conteúdo de vídeo, e é sabido que esta informação tem uma influência direta sobre a forma como o vídeo resiste à perda de pacotes. Embora o mecanismo ACFEC pareça ser uma boa solução quando a rede é estável e há pouca perda de pacotes, quando ocorre congestionamento da rede, este mecanismo gera muito mais pacotes de redundância FEC, que aumentam ainda mais o congestionamento.

Outra técnica para melhorar a qualidade da transmissão de vídeo é feita por meio de *forward error correction* e com retransmissão adaptável baseada em controle de

taxa do canal fonte [Hassan and Landolsi 2010]. Este esquema utiliza monitoramento em tempo real da ocupação do *buffer* do decodificador e o estado do canal para calcular os parâmetros da redundância do FEC. Esta informação é o *feedback* para o codificador do vídeo no servidor, que procede com a adaptação de seus próprios parâmetros de transmissão. Embora os autores afirmem que há uma melhoria no QoE para os usuários finais, o principal objetivo deste esquema é garantir a continuidade da reprodução de vídeo sob variações imprevisíveis de canal e evitar a redundância de FEC desnecessária. Informações como conteúdo de vídeo e tipo de quadro não são consideradas na definição proposta [Aguiar et al. 2012]. Esta abordagem não usa avaliações por métricas de QoE, pois se baseia em valores de perda de pacotes para predizer os níveis de QoE, além de não medir a sobrecarga introduzida.

Outras propostas para melhoria da transmissão de vídeo através de redes locais sem fio baseiam-se num método que adapta, em tempo real, a quantidade de redundância de FEC e a taxa de transmissão [Alay et al. 2010]. Os receptores periodicamente enviam pacotes de informação da taxa de erros para o Ponto de Acesso (AP). Usando essas informações, o AP pode identificar a condição do canal e em seguida, ajustar a taxa de transmissão e FEC. A redundância de FEC é dada por vários vídeos pré-codificados com diferentes taxas de *bitrate* e FEC, assim, a fim de adaptar estes parâmetros o sistema tem de mudar para um fluxo de bits diferente. A necessidade de vídeos pré-codificados reduz a aplicabilidade desta solução. Ela também exige alto poder de processamento e espaço de armazenamento, uma vez que existe a necessidade de codificar várias vezes o mesmo vídeo com taxas de bits e redundância de FEC diferentes. Além disso, só a quantidade de sobrecarga causada pelo FEC introduzido por este mecanismo foi de 48% (sem levar em conta a sobrecarga causada por mensagens de *feedback*, o que é cerca de 6% maior do que o FECAV3D, proposto nesse trabalho, tal como será evidenciado no capítulo 6.

Uma proposta adicional usa a transmissão simultânea por multi-rotas com *path interleaving* para melhorar a transmissão de vídeo [Tsai et al. 2011]. As técnicas mencionadas acima são combinadas com blocos de FEC de tamanho dinâmico. O tamanho do bloco FEC é adaptado de acordo com a taxa média de perda de pacotes para cada caminho, o que permite o envio simultâneo de dados com *path-interleaving* e FEC, ao longo de múltiplas rotas. Esta solução é baseada em parâmetros de rede e não usa características de vídeo, como codec, tipos de quadros, tamanho do GOP, movimentação e complexidade. Além disso, o mecanismo utiliza um *buffer* para lidar com o impacto da desordem de pacotes, devido à transmissão por múltiplas rotas. Isto aumenta o atraso e leva à rejeição dos pacotes pelo decodificador devido ao tempo limite de reprodução. Da mesma maneira como mencionado acima, a sobrecarga de rede também não foi avaliada.

Outra técnica para melhoria da qualidade da transmissão de vídeo é realizada por meio de um esquema baseado em FEC [Díaz et al. 2011]. Esta abordagem utiliza uma proteção desigual de erros para garantir a entrega das informações mais importantes. Os parâmetros que definem a importância do conjunto de dados são a relevância dos quadros, o estado do canal e a taxa de bits. Com estas informações, o algoritmo proposto é capaz de selecionar as imagens mais adequadas para proporcionar uma proteção em tempo real.

Esta solução é baseada apenas no tipo de quadro e não leva em conta a intensidade das atividades espaciais em sequências de vídeo, que podem ter uma influência considerável nas deficiências percebidas. Esta proposta também requer a adição de *tags* extras no encapsulamento do RTP, o que significa que os vídeos são reprocessados.

Apesar da aplicabilidade dentro de vídeos 2D, nenhum dos trabalhos mencionados previamente considera fluxos e tráfego de vídeos 3D. Em [Chung 2010] e [Jean et al. 2012] são propostos mecanismos de Descarte Seletivo de Pacotes para otimização de QoE em vídeos 3D. Em [Hewage et al. 2010] é proposto um estudo de priorização para distribuição de vídeos 3D. Estes trabalhos foram concebidos para funcionar com a representação *video-plus-depth*, onde é atribuída uma prioridade mais elevada à camada base (mapa de cores) em detrimento da camada de reforço (mapa de profundidade), de modo a melhorar a qualidade percebida de vídeo 2D e 3D. Neste último trabalho, fluxos de dados da camada base são mapeados para classes de acesso com maior prioridade que diferem em termos de QoS (perda de pacotes e latência). Embora seja dada prioridade à camada de base, as abordagens propostas não levam em consideração conteúdo do vídeo e os resultados são pouco discutidos e analisados.

Em [Aksay et al. 2009] é estudado o efeito de um mecanismo baseado em FEC para a melhora de tráfegos 3D em redes para TV digital em diferentes condições de canal. Neste trabalho os autores concluem que abordagens que utilizam UEP conseguem alcançar níveis bem melhores de Qualidade de Experiência do que abordagens que não fazem diferenciação dos pacotes de vídeo. Por estas razões, foi proposto e validado o mecanismo FECAV3D, que melhora a qualidade de transmissão de vídeo sem adicionar a sobrecarga da rede desnecessária.

### 4.3 Conclusões

Após um estudo do estado-da-arte é possível concluir que as soluções existentes não levam em consideração QoE e conteúdo do vídeo ou não fornecem uma análise da sobrecarga gerada na rede. Além disso, a avaliação do desempenho da maioria das soluções é deficiente e não verifica o impacto real do ponto de vista do usuário.

Portanto, no próximo capítulo um novo mecanismo é proposto e as avaliações de QoE para o mesmo são apresentadas no Capítulo 6. Os pontos fortes e fracos do FECAV3D são também discutidos no Capítulo 6.

---

---

## CAPÍTULO 5

---

# MECANISMO FEC ADAPTATIVO PARA VÍDEO 3D (FECAV3D)

O aumento das formas de criação, armazenamento e compartilhamento de conteúdo multimídia mostra que um crescimento exponencial do tráfego de vídeo pode ser esperado nos próximos anos [Index 2012]. Neste contexto, o vídeo 3D está previsto a atrair mais e mais o mercado multimídia com a perspectiva de aplicações avançadas (vigilância, controle de missões críticas, entretenimento, etc.) No entanto, as demandas de largura de banda aumentaram em função do conteúdo 3D, bem como os requisitos de Qualidade de Serviço (QoS) e Qualidade de Experiência (QoE).

Fornecer um bom suporte de rede para este tipo de tráfego requer o desenvolvimento de mecanismos adaptáveis que levem em conta as características dos vídeos que serão transmitidos. Através do desenvolvimento destes esquemas, é possível otimizar a transmissão de vídeo 3D assegurando melhor QoE. Técnicas de otimização devem ser utilizadas para melhorar a configuração destes esquemas baseados em FEC, o que lhes permite reforçar as transmissões de vídeo, enquanto aumenta-se a satisfação do utilizador e melhora-se a utilização dos recursos sem fio.

### 5.1 objetivos

Será apresentado uma visão geral sobre o mecanismo FEC Adaptativo para Vídeo 3D (FECAV3D). Este mecanismo pode ser utilizado em diversos tipos de rede e o seu objetivo principal é o de reduzir a sobrecarga da rede, mantendo uma boa qualidade de vídeo. Após isso, no próximo capítulo, este mecanismo será aplicado, testado e analisado dentro de WMNs.

O Mecanismo proposto tem como objetivo: melhorar a qualidade da transmissão de vídeo 3D através da configuração dinâmica do FEC baseado em parâmetros de vídeo, permitindo QoE aceitável e transmissão sem a adição de sobrecarga desnecessária na rede, causada por esquemas FEC típicos. A configuração dinâmica é realizada de acordo com as características de vídeo e da qualidade percebida pelo utilizador, e pode permitir que a informação redundante seja adicionada apenas às partes mais importantes do vídeo, o que significa que a parte do vídeo (que teria um impacto maior sobre a qualidade percebida, se for perdida) está protegida.

## 5.2 Visão geral do Mecanismo FECAV3D

Neste ponto, tem-se uma visão geral do mecanismo FEC Adaptativo para Vídeo 3D (FECAV3D) com esquema de Proteção Desigual de Erros (UEP) para melhorar a transmissão de vídeo 3D. Como mencionado anteriormente, o objetivo é reduzir o *overhead* imposto pelos mecanismos FEC existentes, mantendo os vídeos com melhores níveis de qualidade e permitindo uma melhor utilização dos canais sem fio. Com o uso do FECAV3D, as decisões são tomadas na camada de rede recorrendo a um módulo chamado Base de Classificação de Vídeo (BCV). A tomada de decisões na camada de rede proporciona uma flexibilidade melhor de implementação, porque o mecanismo FECAV3D pode ser implementado em pontos de acesso, *routers* ou no servidor de vídeo. Através da análise das informações coletadas a partir desse módulo, o mecanismo proposto é capaz de estimar a razão de redundância ideal necessária para manter uma boa qualidade de vídeo, sem adicionar sobrecarga desnecessária de rede.

A Fig. 11 mostra o mecanismo FECAV3D. Inicialmente, este usa um classificador de vídeo (BCV) que busca informações, afim de identificar características de vídeo como níveis de movimentação, complexidade espaço-temporal, número de camadas e tamanho do GoP para cada camada. Após isso, é realizado um *Deep Packet Inspection* (DPI) para filtrar informações como o tipo, camada, posição relativa de cada quadro dentro de cada GoP e nível de atividade espaço/temporal através do tamanho do pacote. Finalmente, os blocos FEC são construídos e uma taxa de redundância dada por UEP é atribuída a cada um. Uma explicação mais detalhada de cada etapa será detalhada adiante.

O procedimento de *Deep Packet Inspection* é importante pois possui funções que podem diferenciar a quantidade de redundância necessária para manter uma boa qualidade de vídeo. A primeira função identifica o tamanho do GoP. Conforme apresentado anteriormente, para sequências de vídeo com longos tamanhos de GoP, a perda de quadros I será mais perceptível pelo usuário final do que em GoPs de tamanho menores. Isso acontece porque o comprometimento do vídeo demorará mais tempo, até a chegada de um novo quadro I que irá corrigir o erro, desta forma, o quadro I precisa de mais redundância. A outra função é usada para identificar o tipo de quadro. Isto é importante porque um quadro I terá mais pacotes redundantes, uma vez que em caso de perda, o impacto na qualidade de vídeo será maior do que a perda da estrutura restante, por exemplo, de um



Tabela 2: Descrição dos vídeos originais

Vídeos para a base	Descrição
Aranha e Abelha	Baixa movimentação e complexidade, aranha comendo uma abelha;
Castelo	Média textura, mudanças de profundidade temporais, baixa movimentação;
Rio	Alta textura, baixa movimentação e complexidade, botes em movimento;
Cachoeira	Alta textura, baixa movimentação e complexidade, água caindo constantemente;
Helicóptero	Media textura, movimentação circular, baixo efeito de profundidade;
Corrida de carros	Alta movimentação, movimento de câmera, cortes em cenas, corrida de carro;
Malabarista	Alta complexidade espacial, movimentação de câmera malabarista em trabalho;
Rua de pedestres	Perspectiva linear, alta movimentação e complexidade, câmera fixa, pessoas se locomovendo;
<i>Breakdancers</i>	Alta textura, alta movimentação, média complexidade, dançarinos de <i>street dance</i> em rápidos movimentos;
Basquete	Alta textura, alta movimentação e complexidade, movimentação de câmera, Jogo de basquete;

e complexidade.

A seleção de sequências de vídeo foi realizada de acordo com as recomendações do *Video Quality Experts Group* (VQEG) [Staelens et al. 2011] e International Telecommunication Union (ITU-T) [Winkler 2009]. Ao longo dos experimentos, 20 vídeos foram avaliados. Dez vídeos foram usados para montar a base de dados e um conjunto diferente de outros dez vídeos foram utilizados para avaliar o mecanismo FECAV3D. Estes vídeos possuem distorções e conteúdos diferentes, sendo portanto bastante representativos como materiais de visualização comuns. Além disso, essas sequências incluem diferentes cores, luminância, trocas de cenas, movimentação e detalhes espaciais. A Tabela 2 apresenta os vídeos utilizados na criação da base de vídeos e uma breve descrição dos mesmos.

Alguns autores tendem a classificar a movimentação e complexidade de vídeo em três ou duas categorias, nomeadamente complexidade baixa, média e alta [Khan et al. 2009] ou complexidade alta e baixa [Aguiar et al. 2012]. Neste trabalho, usaou-se a classificação de duas categorias (alta e baixa complexidade) de maneira a simplificar o mecanismo. Esta classificação é apresentada através dos clusters do dendrograma da fig. 12, com a *Linkage Distance* (LD) 2.

O comportamento destes dois grupos pode ser observado nos exemplos mostrados na fig. 13 e Fig. 14, que descrevem duas sequências de vídeo - Breakdancers e Bullinguer, respectivamente - cada uma de um conjunto diferente. Para melhor visualizar os resultados, apenas o primeiro GOP de cada vídeo e camada foi considerado (mais informações

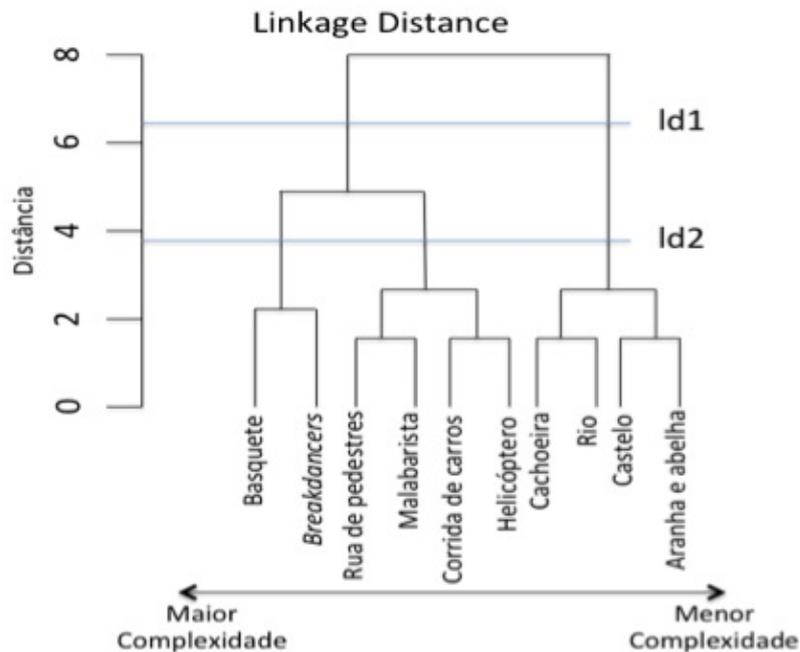


Figura 12: Árvore das distâncias euclidianas

sobre as características do codec de vídeo podem ser encontradas no Capítulo 2). O vídeo Breakdancers tem movimentação contínua na cena e uma câmera de ângulo amplo, portanto, alta movimentação e complexidade. Desta forma, este vídeo tem quadros maiores e maior diferença de tamanho entre quadros P e B, como ilustrado pela Fig. 13. A seqüência Bullinguer tem uma pequena região de movimento e de interesse, apenas o rosto, e um fundo estático. Por conseguinte, mostra baixa movimentação e complexidade e há uma diferença menor de tamanho entre quadros P e B, como mostrado na fig. 14.

As Figs. 13 e 14 também ilustram os resultados das avaliações do Índice de Semelhança Estrutural (SSIM - *Structural Similarity Index*) quando os quadros são intencionalmente removidos do GOP. A medição dessa métrica é bastante simples, no entanto, é bastante consistente com o sistema visual humano, dando bons resultados [Wang et al. 2004a]. Os valores de SSIM foram obtidos através da remoção do quadro que ocupava uma determinada posição, ou seja, o valor calculado do primeiro SSIM foi feito sem o primeiro quadro, o valor calculado do segundo SSIM foi feito sem o segundo quadro e assim por diante. Através da análise destes resultados pode-se perceber que, na seqüência Breakdancers, além do fato dos quadros I e P serem os mais importantes, os quadros mais próximos do início do GOP tem um maior impacto na qualidade do vídeo quando removidos, tanto na camada base quanto na camada de profundidade. Por outro lado, a seqüência Bullinguer tem um comportamento um pouco diferente, devido à sua menor movimentação e complexidade, o que aumenta a capacidade de resiliência à perda de pacotes [Khan et al. 2009].

Utilizando as informações mencionadas acima, a saída do módulo BCV identifica os níveis de movimentação e complexidade de cada GOP que está sendo transmitido assim como o tamanho do GOP. Apesar do tamanho do GOP permanecer o mesmo, estes

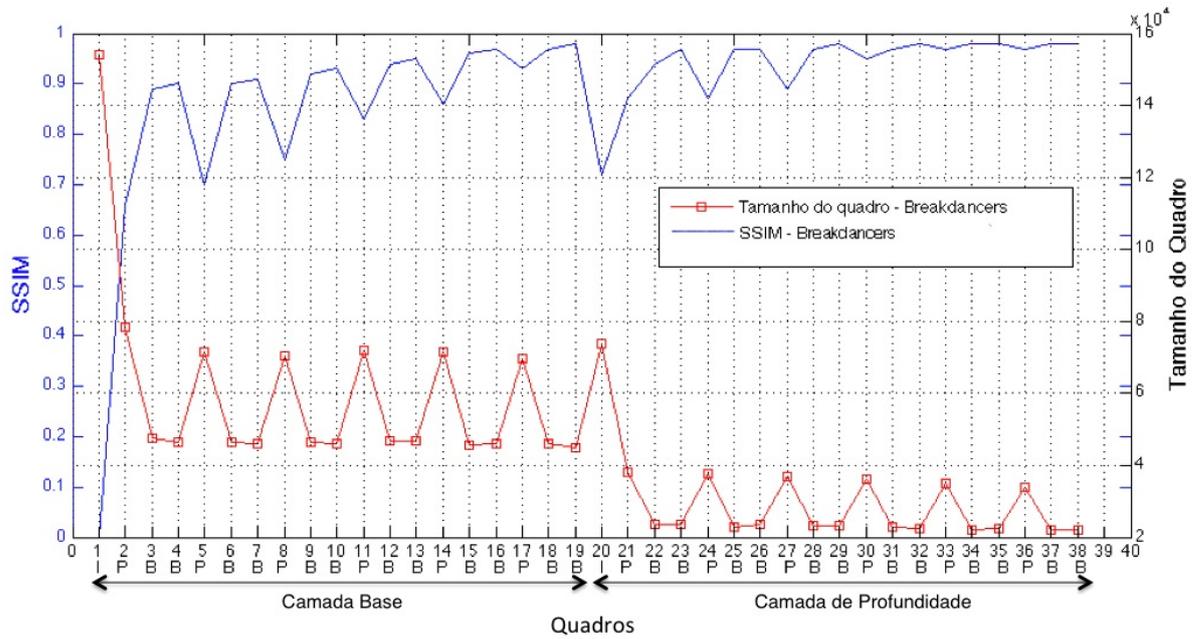


Figura 13: Breakdancers - Tamanho dos quadros x QoE (SSIM)

parâmetros são definidos GOP a GOP (Fig. 11), uma vez que é possível ter diferentes níveis de movimentação e complexidade dentro de uma mesma sequência de vídeo, tal como esperado para vídeos da Internet. Após isso, a informação sobre o tipo de quadro e a posição relativa no interior do GOP (para quadros P), é obtida a partir de DPI. Esta informação é importante para identificar as características dos vídeos que serão necessárias na fase seguinte, para configurar a quantidade de redundância.

Finalmente, uma quantidade pré-definida de redundância é utilizada para configurar o esquema de FEC. O mecanismo FECAV3D tem uma estrutura modular que permite ser alterado se necessário. Durante os experimentos, o código *Reed-Solomon* (RS), foi utilizado. Este código oferece menor complexidade, e, portanto, melhor desempenho para serviços em tempo real [Neckebroek et al. 2010]. Um código de RS é composto pelos elementos  $\mathbf{n}$ ,  $\mathbf{s}$  e  $\mathbf{h}$ , conforme ilustrado na Fig. 15. O tamanho do bloco total, incluindo os dados de redundância, é representado por  $\mathbf{n}$ , e  $\mathbf{s}$  indica o tamanho do conjunto original de dados, assim, o código de paridade é  $(\mathbf{n}, \mathbf{s})$ . O último parâmetro é  $\mathbf{h}$  e define a quantidade de redundância ( $\mathbf{h} = \mathbf{n} - \mathbf{s}$ ). A fim de recuperar todos os dados originais ( $\mathbf{s}$ ), pelo menos,  $\mathbf{n} - \mathbf{h}$  pacotes devem ser enviados com sucesso. A robustez à perdas é determinado pelo tamanho do  $\mathbf{h}$ , e a capacidade de recuperação de erros contra uma taxa média de perda de pacotes pode ser expressa como  $\mathbf{h} / \mathbf{n}$  ou  $(\mathbf{n} - \mathbf{s}) / \mathbf{n}$ .

No mecanismo FECAV3D o código de paridade é ajustado em tempo real. Isto significa que os parâmetros  $\mathbf{n}$  e  $\mathbf{h}$  são determinados no fim, de acordo com as características de vídeo, obtidos previamente a partir do módulo BCV e DPI. O parâmetro anterior  $\mathbf{n}$ , é usado para construir um Bloco Flexível de FEC (BFF), e o último parâmetro  $\mathbf{h}$  contém a quantidade adaptada de redundância para cada BFF. O BFF é a divisão de quadros I e P em grupos de pacotes. Cada grupo pode ter uma quantidade diferente de redundância de dados. Por este motivo, em vez da adição de uma taxa única de

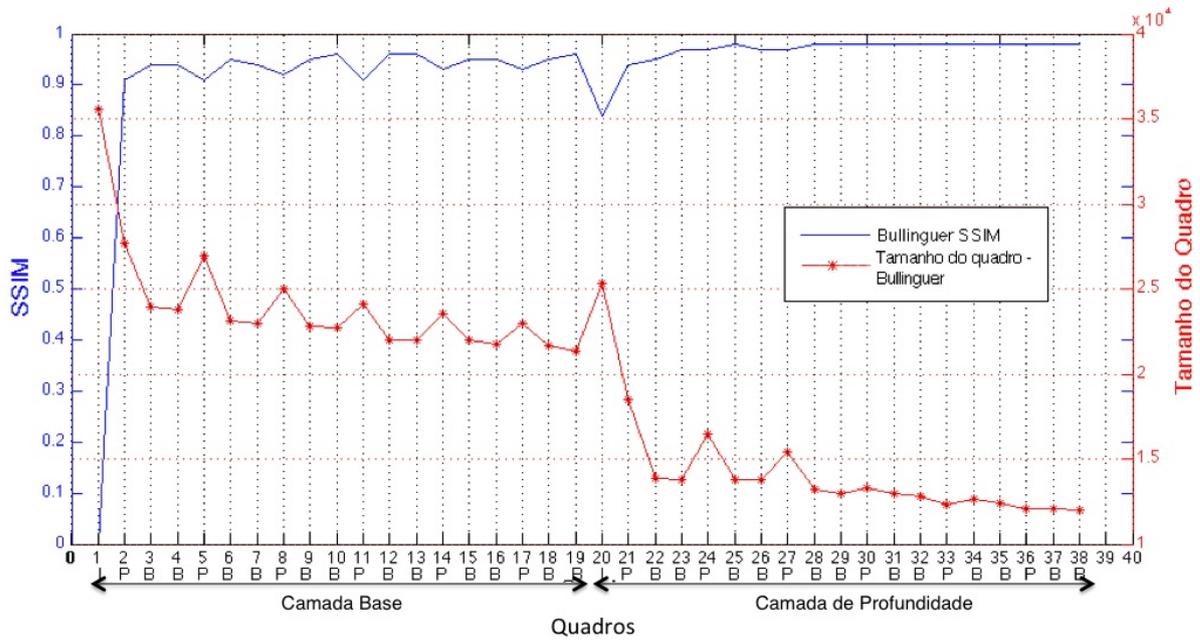


Figura 14: Bullinger - Tamanho dos quadros x QoE (SSIM)

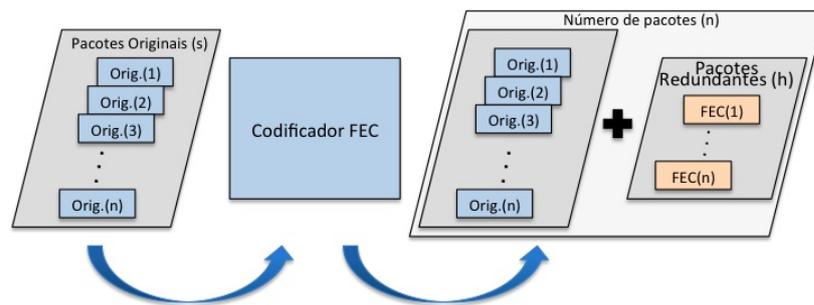


Figura 15: Parâmetros de Correção Antecipada de Erros (FEC)

redundância a esses quadros, e conseqüentemente para as sequências de vídeo, constrói-se uma estrutura flexível, a fim de se ajustar para vários tipos diferentes de vídeo e condições de rede.

A quantidade de redundância de dados definida pelo FECAV3D é dada como a combinação da movimentação e complexidade de vídeo, posição do quadro no GOP e tipo de quadro. Com esta combinação permite-se inferir características espaço-temporais de vídeo, e, portanto, escolher a quantidade de redundância ideal  $h$  para cada BFF. Ao fazer isso, o FECAV3D atinge uma melhor qualidade de vídeo, e também reduz a quantidade de dados que deve ser enviada através da rede, diminuindo a sobrecarga. A redução da sobrecarga da rede é muito importante porque, quando a rede torna-se maior, há um aumento no número de transmissões simultâneas, causando graves problemas de interferência. Esta situação fica ainda pior se somarmos mais sobrecarga devido a informações de redundância. Portanto, se conseguirmos reduzir a sobrecarga, mais usuários serão capazes de receber mais vídeos com melhor qualidade, melhorando o desempenho geral da rede.

Para ilustrar o funcionamento do mecanismo FECAV3D, a Fig. 16 mostra o pseudocódigo para a detecção do tamanho do GOP e dos níveis de movimentação, bem como os passos da atribuição de um valor de redundância. Todas as operações são feitas dentro de dois *loops*. O primeiro, na linha 1, vai passar por todos os GOPs em uma sequência de vídeo. O segundo, na linha 4, está dentro do primeiro e irá percorrer todos os quadros dentro de um GOP, aplicando apenas a redundância necessária. Nas linhas 2 e 3 é possível notar o acesso ao módulo BCV e nas linhas 5, 7, 12 e 13 é utilizado DPI. Existe uma diferença importante no tratamento de quadros I e P : a camada em que estão inseridos e a posição do quadro P. Desta forma, como é possível ver na linha 13, 7 e 12, a quantidade de redundância do quadro P depende também da sua posição relativa dentro do GOP e camada que ocupa.

```

01 for each GoP
02   BCV.getGopMotion(GoP)
03   BCV.getGopSize(GoP)
04   for each frame
05     case (DPI.getFrameType(frame))
06       I-Frame:
07         DPI.getLayer(frame)
08         buildBFF(frame)
09         addRedundancy(frame)
10         sendFrame(frame)
11       P-Frame:
12         DPI.getLayer(frame)
13         DPI.getRelativePosition(frame)
14         buildBFF(frame)
15         addRedundancy(frame)
16         sendFrame(frame)
17       B-Frame:
18         sendFrame(frame)
19     end case
20   end for
21 end for

```

Figura 16: Pseudocódigo para o FECAV3D

A quantidade de redundância adicionada pelo mecanismo FECAV3D em cada GOP ( $R_{GoP}$ ) pode ser calculada como é apresentado em 5.1.  $FS_i$  representa o número de pacotes do quadro que está sendo transmitido e  $FT_i$  representa o tipo de quadro como descrito em 5.2. Se  $\gamma > 0$ , isso significa que um certo nível de redundância será adicionado ao quadro. Por exemplo, um vetor (1, 1, 0) significa que quadros I e P receberão redundância, mas não quadros B. Uma configuração adicional poderia ser (2, 1, 0), que indica que os quadros I devem receber o dobro da redundância que é fornecida para quadros P (se os outros parâmetros são os mesmos). Esta configuração pode ser usada se existe uma necessidade de se melhorar ainda mais a qualidade de vídeo, mesmo que para isto, leve a um aumento da sobrecarga da rede.

Os níveis de movimentação e complexidade são descritos por  $C_{GoP}$  como visto em 5.3. Se o mecanismo está usando dois *clusters* distintos de vídeos, os valores podem ser, por exemplo, (1, 0.5). Desta forma, o *cluster* com elevado nível de movimentação e complexidade recebe o dobro da redundância do *cluster* com nível baixo. Por outro lado, se houver uma necessidade de utilizar mais níveis, outra opção seria a utilização de (1, 0.5, 0.25), o que significa que três níveis de movimento e de complexidade serão tratados, alto, médio e baixo, respectivamente. Os dois últimos parâmetros em 5.1 são  $RP_i$  e  $FL_i$ , definindo a distância relativa de quadros P dentro do GOP e a camada do

Tabela 3: Tabela de notações

Notação	Significado
$R_{GoP}$	Quantidade de redundância por GoP
$FS_i$	Tamanho do quadro em pacotes
$FT_i$	Tipo do quadro
$C_{GoP}$	Nível de movimentação e complexidade do GoP
$RP_i$	Posição relativa do quadro P
$FL_i$	Camada do quadro
$N_{GoP}$	Número de GoPs no vídeo

quadro, respectivamente. Quadros mais perto do fim do GOP receberão uma quantidade reduzida de redundância porque o impacto de perda de pacote será menor do que a perda perto do início do GOP. Quadros da camada de profundidade recebem menor prioridade e menor taxa de redundância também, pois se há perdas na camada base, ainda se pode visualizar um vídeo 2D. A Tabela 3 apresenta a notação usada nas Equações.

$$R_{GoP} = \sum_{i=0}^{GoPSize} \left[ FS_i \times FT_i \times C_{GoP} \times \frac{1}{RP_i} \times \frac{1}{FL_i} \right] \quad (5.1)$$

$$FT_i = \begin{cases} \gamma > 0 & , \text{envia quadro com redundância} \\ 0 & , \text{quadro sem redundância} \end{cases} \quad (5.2)$$

$$C_{GoP} = \begin{cases} 1 & , \text{alta moviment. e complexidade} \\ 0 \leq \alpha < 1 & , \text{outros casos} \end{cases} \quad (5.3)$$

A quantidade total de redundância dentro de uma sequencia de vídeo é a soma de cada redundancia de GoP ( $R_{GoP}$ ). Onde é possível encontrar a quantidade média de redundância ( $R$ ) através de 5.4.

$$R = \frac{1}{N_{GoP}} \sum_{i=0}^{N_{GoP}} R_{GoP(i)} \quad (5.4)$$

### 5.3 Conclusões

Este capítulo apresentou o mecanismo, baseado em FEC/UEP, proposto no trabalho (FECAV3D) que realiza a inserção de pacotes de redundância de acordo com a importância, a camada, o nível de atividade espaço-temporal e a posição dos quadros dentro dos vídeos. Além de melhorar o impacto da qualidade dos vídeos transmitidos,

sob o ponto de vista do usuário final, uma otimização dos recursos da rede sem fio é assegurados.

O mecanismo opera levando em consideração a prioridade de cada quadro e camada dentro do fluxo, e a quantidade de redundância é ajustada para assegurar que ele está em conformidade com as características dos vídeos que estão sendo transmitidos. Este ajuste pró-ativo ocorre em tempo real, e toda a informação é recolhida através de técnicas de DPI com a ajuda de um banco de dados de atividade espaço-temporal que é construído antes da execução do mecanismo. O estabelecimento do banco de dados é realizado apenas uma vez, após o que, pode ser consultado em tempo real. O objetivo principal é obter a melhor qualidade de vídeo possível sem adicionar sobrecarga de rede desnecessária, devido aos pacotes de redundância. A avaliação desses mecanismos, bem como, o cenário de simulação e os resultados obtidos são descritos no Capítulo 6.

---

---

## CAPÍTULO 6

---

# SIMULAÇÕES E AVALIAÇÕES

O principal objetivo do mecanismo FECAV3D é o de reduzir a sobrecarga da rede introduzida pelos esquemas baseados em FEC padrões, mantendo os vídeos 3D em um nível de qualidade aceitável. A fim de avaliar os benefícios e impactos do FECAV3D em WMNs, experimentos foram realizados utilizando o *Network Simulator 2* (NS-2) [Weingartner et al. 2009]. O cenário de avaliação é composto de seis nós distribuídos em forma de grade (3x2), cada nó está a 90 metros de distância do vizinho mais próximo [Oh and Chen 2010]. O *Optimized Link State Routing Protocol* (OLSR) [Jacquet 2003] foi utilizado como protocolo de roteamento, no entanto, qualquer outro protocolo pode ser utilizado. Tráfegos com Taxa de Bits Constante (CBR - *Constant Bit Rate*) foram definidos como *background* a uma taxa de 800 kbps e 10 sequências de vídeo 3D foram utilizadas no cenário de avaliação [Fitzek and Reisslein 2001], com tamanho (1024x768), codec H.264/MPEG-4 Parte 10 e 8 Mbps [Lo et al. 2011]. O tamanho do GOP foi definido para 19:2, o que significa que a cada 19 quadros, teremos outro quadro I e depois de cada dois quadros B haverá um quadro P. O decodificador usa o *Frame-Copy* como método de substituição de erros, isto significa que o mesmo substituirá cada quadro perdido com uma cópia do último bem recebido.

Além do tráfego de fundo, um modelo de cadeia de Markov de dois estados foi utilizado para melhor refletir ambientes de rede na prática. Este modelo também é conhecido como modelo de perda Gilbert-Elliot. Ele é usado para a produção de resultados de simulação mais realistas, pois simula padrões de perda comuns em canais sem fio [Wilhelmsson and Milstein 1999]. Este modelo é composto de dois estados ( $G$ ,  $B$ ) e de quatro parâmetros ( $PG$ ,  $PB$ ,  $r$ ,  $k$ ), a Fig. 17 mostra os parâmetros que são empregados. No estado  $G$  (“Good” ou “Gap”) as mensagens são entregues e o parâmetro  $PG$  possui uma probabilidade que indica que um pacote chegou com sucesso e sem quaisquer tipos de erros. O inverso é verdadeiro para o estado  $B$  (“Bad” ou “Burst”), e o parâmetro  $PB$ , que indica a probabilidade de um pacote ser perdido ou danificado. O estado de transição

é dado por  $k$  e  $r$ , onde  $k$  representa a probabilidade de transição do estado  $G$  para  $B$  e  $r$  representa a probabilidade de transição do estado  $B$  para  $G$ .

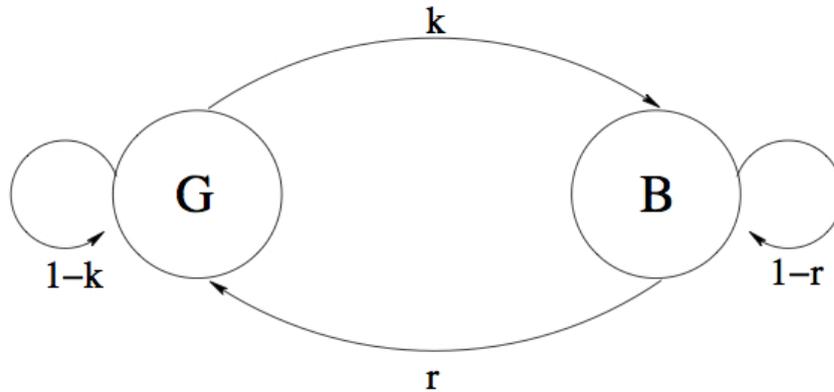


Figura 17: Modelo de Markov com 2 estados

É possível calcular a probabilidade de estado estacionário para os estados  $G$  e  $B$ ,  $\varphi_G$  e  $\varphi_B$  respectivamente, com as Equações 6.1 e 6.2.

$$\varphi_G = \frac{r}{r + k} \quad (6.1)$$

$$\varphi_B = \frac{k}{r + k} \quad (6.2)$$

Também é possível calcular a probabilidade de perda de pacotes,  $P_{loss}$ , usando a Equação 6.3.

$$P_{loss} = P_G \times \varphi_G \times P_B \times \varphi_B \quad (6.3)$$

Foram realizados três experimentos com esquemas diferentes a fim de comparar os resultados. No primeiro experimento, o qual serviu como uma "baseline", foi realizado sem qualquer mecanismo (sem FEC). O segundo foi implementado com uma abordagem FEC simples e não adaptativa (FEC padrão), onde uma quantidade fixa de redundância de dados (80%) foi estaticamente adicionada a ambos, I e P-frames. A quantidade de redundância de dados foi definida com base em uma série de experiências exaustivas, que mostraram, de acordo com as condições da rede, a melhor qualidade de vídeo 3D recebida. Finalmente, o último experimento foi realizado utilizando a abordagem proposta adaptativa com Proteção Desigual de Erros (FECAV3D). Cada um destes três experimentos foi simulado 10 vezes com diferentes porcentagens de perda devido a diferentes *seeds* iniciais para a geração de números aleatórios utilizados pelo modelo Gilbert-Elliot [Salyers et al. 2008]. A perda média foi de aproximadamente 20%.

A qualidade de cada vídeo 3D obtida na avaliação dos diferentes experimentos foi avaliada através de métricas objetivas e subjetivas. Métricas objetivas utilizam um conjunto de indicadores de qualidade correlacionados à percepção do usuário para realizar a avaliação, eliminando assim, a intervenção humana. Como explicado no capítulo 3, o PSNR não analisa valores de cromaticidade dos *pixels*, não havendo portanto, correlação dos valores de PSNR dos vídeos 3D (devido aos mapas de profundidade) com a percepção subjetiva humana, desta forma, as métricas objetivas utilizadas para avaliação da qualidade do vídeo foram SSIM e VQM. A avaliação da qualidade objetiva das sequências de vídeo 3D foi realizada com Evalsvc [Le et al. 2010] e MSU *Video Quality Measurement Tool* (VQMT) [Vatolin et al. 2009].

Embora testes objetivos possam facilmente avaliar a qualidade do vídeo, estes não conseguem captar todos os detalhes que podem afetar a experiência do usuário, e, portanto, avaliações subjetivas também são necessárias, neste caso, o MOS com SS e ACR foram utilizados. O MOS é um dos métodos mais utilizados para a avaliação subjetiva dos vídeos 3D e 2D. É recomendado pela ITU-T e utiliza um conjunto de pessoas que votam em sequências de vídeo, de acordo com uma escala de qualidade pré-definida, para avaliar a qualidade. A escala MOS vai de 1 a 5, em que 5 é o melhor resultado possível conforme retratado previamente no capítulo 3. A norma ITU-R BT.500-11 [ITUR-B500-11 2002] com o método *Single Stimulus* (SS) foi utilizada na avaliação subjetiva deste trabalho. Neste método, os telespectadores assistem cada vídeo 3D apenas uma vez e, em seguida, avaliam a qualidade usando a escala *Absolute Category Rating* (ACR) (Ruim; Pobre; Médio; Bom e Excelente), que está associada à escala MOS. O paradigma SS foi escolhido porque é adequado para realizar a avaliação da qualidade de aplicações emergentes de vídeo [Seshadrinathan et al. 2010].

Os testes subjetivos foram realizadas na Universidade Federal do Pará em um laboratório especialmente equipado para satisfazer as recomendações feitas no ITU-R BT.500-11 para avaliação em ambientes domésticos. Foi desenvolvido um aplicativo para exibir as sequências de vídeo 3D e recolher as pontuações dos usuários. A tela usada para apresentar as sequências foi uma televisão 46" SAMSUNG estereoscópica, com 1920x1080 de resolução. Para visualizar as sequências em 3D, os observadores usavam óculos obturadores ativos em conjunto com este aparelho de TV. A distância entre os olhos dos participantes e a TV foi fixada como sendo cinco vezes a altura da sequência de vídeo, como recomendado pela ITU-T. As cadeiras também foram ajustadas de modo que os olhos dos participantes ficassem alinhados com o mesmo nível horizontal e vertical do centro da TV. Todas as sequências foram reproduzidas em ordem aleatória na tela. 30 observadores participaram dos experimentos, tendo os mesmos, visão normal e idade variando de 18 a 45 anos de idade. Foi assegurado que eles passassem no teste de daltonismo (teste de Ishihara) e no teste de percepção de profundidade (teste de Randot). Os observadores incluíram estudantes de graduação, pós-graduação e funcionários da universidade.

A Figura 18 mostra os resultados dos experimentos subjetivos. Sem o uso de um esquema baseado em FEC para proteger a transmissão, a média do MOS foi de 2.04, que é considerado vídeo 3D com má qualidade e deficiências bastante incômodas.

Quando o FEC simples e não adaptativo (FEC padrão) e o mecanismo FECaV3D foram empregados, os valores médios de MOS foram de 4,19 e 4,16, respectivamente. Estes valores estão entre boa e excelente qualidade, com deficiências perceptíveis, mas não intoleráveis. Os resultados mostraram que um dos objetivos do mecanismo FECaV3D foi alcançado, que é de manter a qualidade do vídeo 3D em melhores condições.

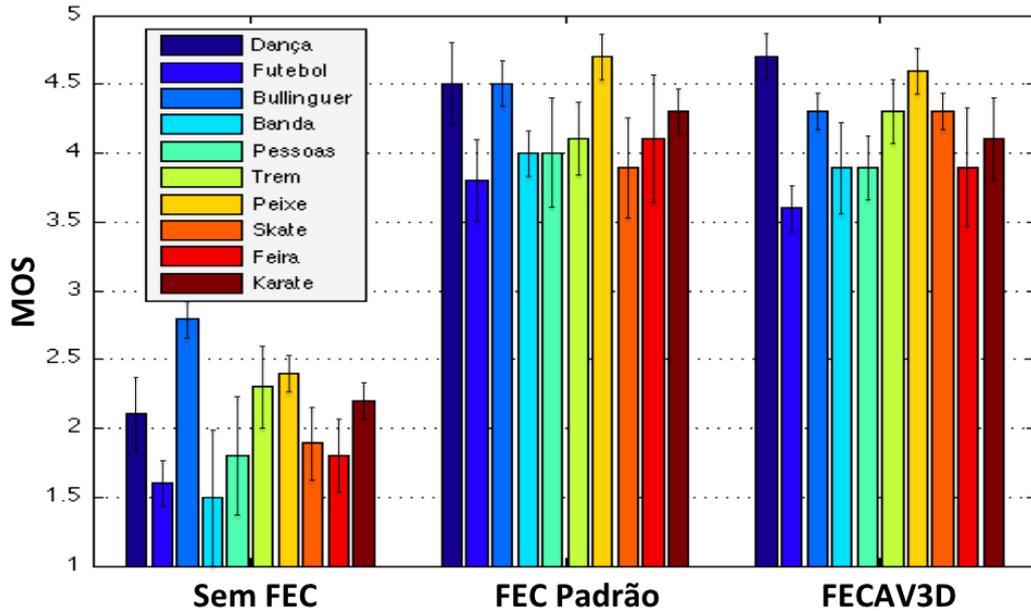


Figura 18: Média do MOS por Vídeo 3D

Métricas objetivas também foram empregadas para confirmar os resultados encontrados através das medições subjetivas. As Figs. 19 e 20 mostram a média dos valores de SSIM e VQM para todas as sequências de vídeos 3D. Na métrica SSIM, valores mais próximos de 1 (um) indicam melhor qualidade de vídeo. Como é ilustrado na Fig. 19, quando há um aumento na taxa de perda de pacotes, há uma diminuição acentuada da qualidade dos vídeos 3D que estão sendo transmitidos sem qualquer tipo de mecanismo de proteção. Ao mesmo tempo, sequências de vídeos 3D que usam qualquer tipo de mecanismo baseado em FEC, são capazes de manter uma melhor qualidade.

Um aspecto importante e que merece destaque, é que, com taxas de perdas de pacotes de 5% a 10%, a qualidade das sequências de vídeos 3D sem FEC se mantém aproximada e não sofre uma grande perda. Isto pode ser explicado pela resiliência natural dos vídeos a uma certa quantidade de perda de pacotes. De modo geral, as sequências de vídeos que possuem menor complexidade espacial e temporal são mais resistentes à perda e alcançam melhores resultados na avaliação de QoE. Outras sequências, com maiores taxas de complexidade espacial e temporal, apresentaram resultados mais pobres e apesar da média semelhante, o desvio padrão foi mais elevado, com a taxa de perda de pacotes em 10%. Isto significa que os valores obtidos das avaliações de QoE, são mais distantes uns dos outros. Quase o mesmo padrão é perceptível nos valores de VQM na Fig. 20. Nesta métrica, vídeos 3D com melhor qualidade possuem pontuações mais próximas a zero. Com uma taxa de perda de pacotes de 5% a 10%, os valores de VQM são também próximos uns dos outros. Isto não é tão evidente como na métrica SSIM, pois o VQM

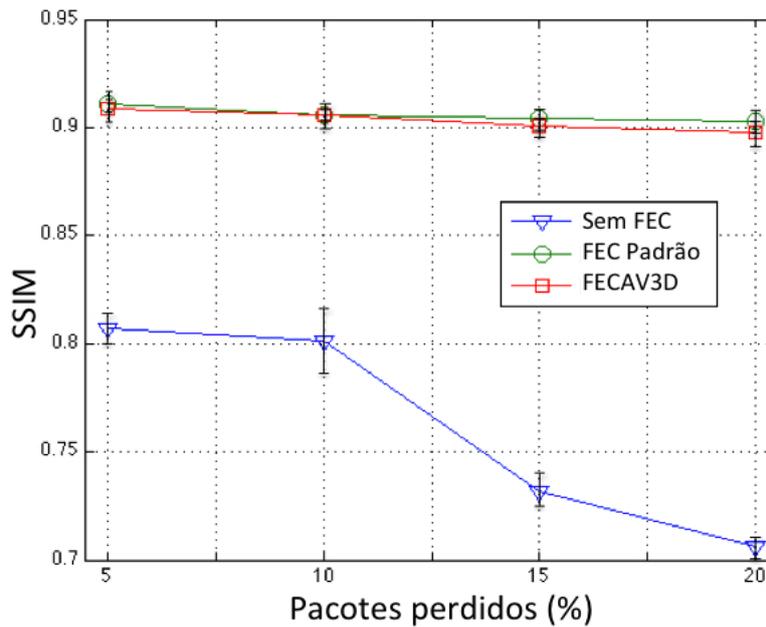


Figura 19: Valores médios de SSIM para todas as sequências de vídeo 3D

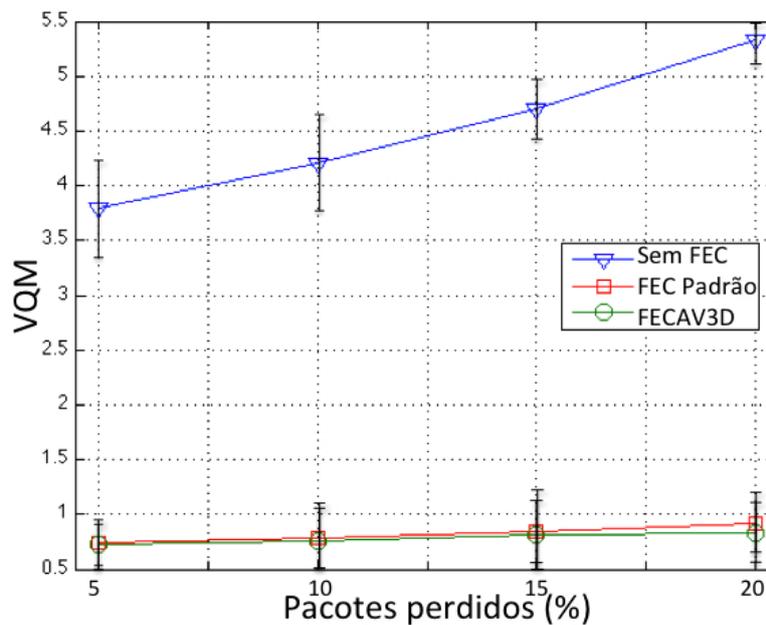


Figura 20: Valores médios de VQM para todas as sequências de vídeo 3D

tende a ser mais rígido em relação a deficiências de vídeo, uma vez que vídeos com maior complexidade espaço-temporal tendem a ter resultados mais pobres. Pela mesma razão, o desvio padrão desta métrica tende a ser maior do que a métrica SSIM.

A Fig. 20 também mostra um grande desvio padrão para o cenário sem FEC. A razão para isto é que a métrica VQM tende a aumentar rapidamente os pontos negativos dos vídeos danificados. Se nenhuma técnica de correção de erros é empregada, alguns vídeos 3D terão mais defeitos e outros menos. Portanto, em função da tendência desta métrica em atribuir pontuações mais baixas para vídeos com maior grau de deficiência;

Tabela 4: Valores de ganho em QoE

Perda de pacotes (%)	Métrica	Sem FEC	FEC Padrão e Ganho	FECAV3D e Ganho
5% de perda	SSIM	0,807	0,911 (12, 89% )	0,909 (12, 64%)
10% de perda	SSIM	0,801	0,906 (13, 11% )	0,906 (13, 11%)
15% de perda	SSIM	0,732	0,904 (23, 50% )	0,901 (23, 09%)
20% de perda	SSIM	0,706	0,903 (27, 90% )	0,898 (27, 20%)
5% de perda	VQM	3,802	0,734 (80, 68%)	0,729 (80, 82%)
10% de perda	VQM	4,251	0,792 (81, 19%)	0,761 (81, 90%)
15% de perda	VQM	4,790	0,840 (82, 13%)	0,828 (82, 55%)
20% de perda	VQM	5,330	0,912 (82, 93% )	0,830 (84, 43%)

vídeos transmitidos sem correção de erros receberão dessa forma notas ruins, resultando em um maior desvio padrão.

A Tabela 4 resume os resultados e mostra os percentuais de ganho de cada cenário. Na métrica VQM, quanto menor a porcentagem, melhor será o resultado. Por outro lado, o oposto também é verdadeiro para a métrica SSIM. Como esperado, os dois mecanismos baseados em FEC produzem resultados melhores (em comparação com o cenário sem FEC) quando a rede tem uma maior taxa de perda de pacotes. Por exemplo, foi possível obter uma redução de mais de 80% nos valores de VQM, o que significa aproximadamente 5 vezes menor pontuação. Com a métrica SSIM, houve um aumento de mais de 25% nos resultados, que significa um melhor controle de QoE. Para as métricas SSIM e VQM, não é possível estabelecer uma correlação linear em função do modelo de análise diferente que cada uma faz, desta forma, os valores obtidos por essas métricas são independentes e não podem ser diretamente comparados [Aguiar et al. 2012].

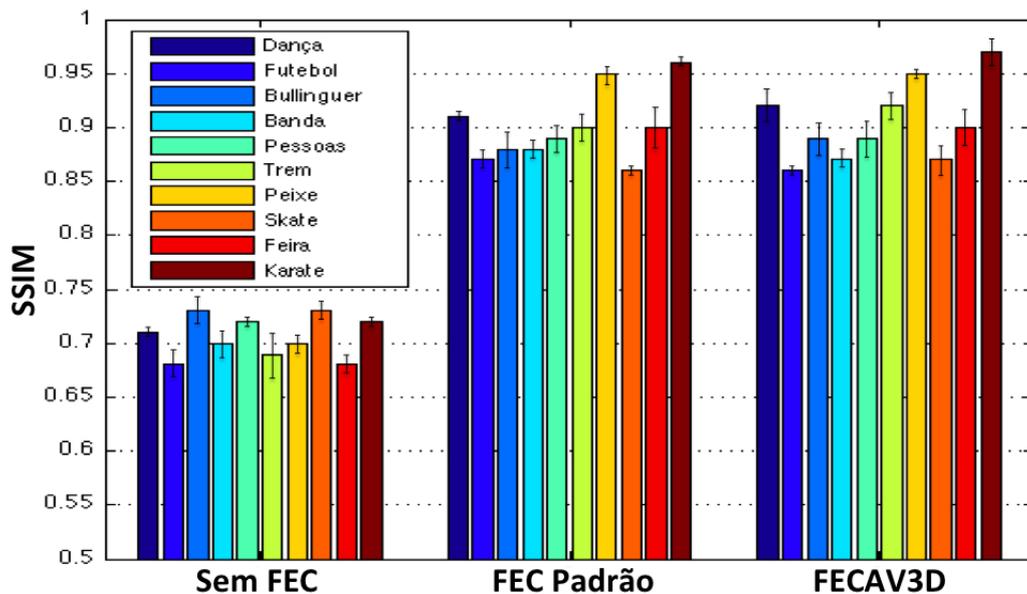


Figura 21: Média do SSIM por Vídeo 3D

Uma análise mais detalhada foi realizada com o melhor cenário (taxa de perda de pacotes de 20%). As Figs. 21 e 22 mostram os valores de SSIM e VQM de cada sequência

de vídeo 3D usada nos testes. O valor médio das métricas SSIM e VQM (quando os mecanismos baseados em FEC não foram utilizados) foi de 0,706 e 5,330, respectivamente. Estes valores representam baixos níveis de qualidade e confirmam as taxas encontradas na avaliação subjetiva. Por outro lado, a média da métrica SSIM do mecanismo baseado em FEC não adaptativo e do FECAV3D foi de 0,903 e 0,898, respectivamente. Os valores médios da métrica VQM foram 0,912 e 0,830, respectivamente. Estas pontuações mostram uma boa qualidade de vídeo 3D, o que mais uma vez, corroboram os resultados subjetivos. As diferenças nos valores de avaliação de vídeo, que podem ser visualizados nos resultados, são, devido às características únicas das sequências de vídeo 3D. Pequenas diferenças nos níveis de complexidade e movimentação podem influenciar os valores obtidos. Em virtude disto, é importante fazer uso de vários tipos de vídeo ao realizar as experiências.

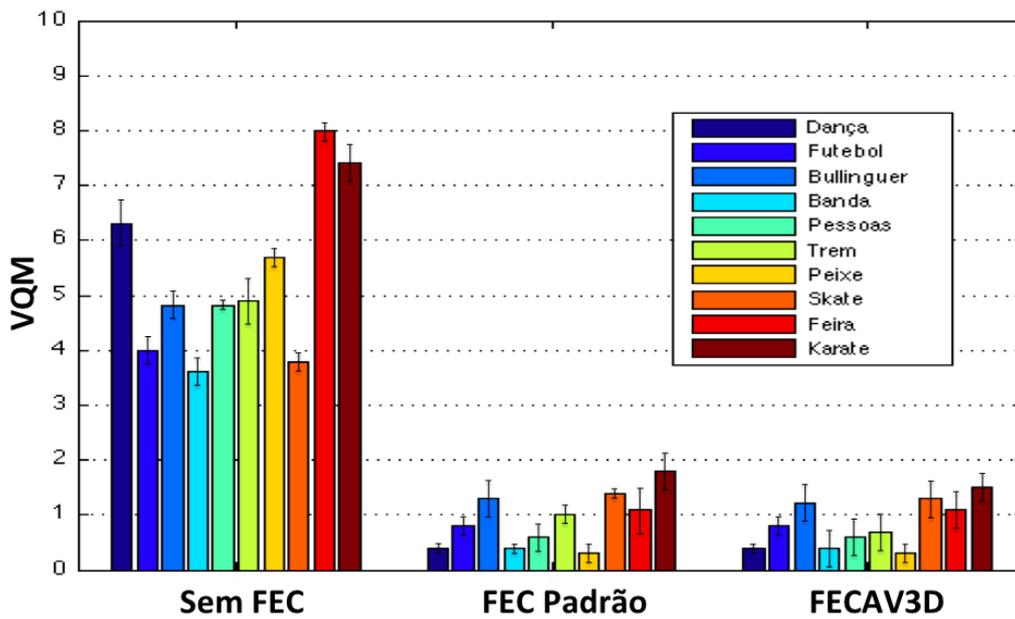


Figura 22: Média do VQM por Vídeo 3D

Todas as avaliações de QoE realizadas, mostraram que o mecanismo FECAV3D foi capaz de manter uma boa qualidade de vídeo 3D. No entanto, o objetivo principal do mecanismo proposto era diminuir a sobrecarga da rede. Isto é importante nas redes sem fio, em especial no caso de WMNs, devido aos recursos limitados do canal sem fio, largura de banda e até mesmo em função da distribuição irregular e das interferências causadas por transmissões simultâneas. Quando o FEC Padrão não adaptativo foi utilizado, a sobrecarga de rede adicional ficou entre 58% e 79%, conforme mostrado na Fig. 23. Por outro lado, quando o mecanismo FECAV3D foi usado, a sobrecarga de rede se manteve entre 36% e 49%. O mecanismo FECAV3D impõe, deste modo, em média, 40% a menos de sobrecarga de rede, com uma qualidade de vídeo igual ou ligeiramente inferior, como mostrado nas Figs. 18, 21 e 22.

O vídeo 3D Skate teve a menor redução na sobrecarga da rede (36,51%) e o vídeo 3D Futebol teve a maior (40,26%). Isto pode, em parte, ser explicado pelo tamanho dos quadros I, P e B que diferem entre estes vídeos, o que gera um número maior ou menor número de pacotes de cada tipo. A Fig. 24 apresenta o número de pacotes de cada tipo

de quadro e camada em todos os vídeos 3D. Os quadros B da camada de profundidade e camada base estão unidos e representados pela cor azul escuro, pois tanto o FECAV3D como o FEC Padrão, não retransmitem quadros do tipo B. Quando os valores de Skate são analisados, percebe-se que mais de 75% dos pacotes pertencem a quadros B, tanto da camada base como da camada de profundidade, e que não são considerados seja no FEC padrão ou no FECAV3D, porque causam danos menores se perdidos. Isto significa que menos de 25% dos pacotes são otimizados pelo mecanismo FECAV3D, e isto resulta em uma menor redução da sobrecarga. Por outro lado, a sequência futebol tem mais de 55% dos pacotes sendo quadros I ou P, o que pode ser otimizado, resultando numa maior redução de sobrecarga da rede.

Uma análise mais aprofundada foi realizada no melhor cenário - vídeo 3D Futebol. Neste caso, além de reduzir a sobrecarga da rede, foi possível obter mais de 65% em média, de melhora nas métricas de SSIM, VQM e MOS, como mostrado nas Figs. 18, 21 e 22, em comparação com o vídeo enviado sem mecanismos para melhorar a qualidade (Sem FEC). Se comparado com o FEC padrão não-adaptativo, O FECAV3D ainda consegue resultados bastante similares com diferença de menos de 2% de melhoria na qualidade de vídeo 3D quando a avaliação da qualidade é realizado por meio da métrica SSIM e chega a empatar, quando a métrica VQM é utilizada. A abordagem subjetiva mostrou também que o FECAV3D obteve resultados bastante próximos. Além disso, o desvio padrão também é pequeno (quando as métricas SSIM e MOS são usadas), o que significa que FECAV3D obtém resultados precisos e indica que se trata de um mecanismo estável e confiável. No que diz respeito à melhoria da qualidade de vídeo 3D e a diminuição da sobrecarga da rede, o FECAV3D também mostrou que obteve uma taxa de recuperação levemente melhor, como pode ser visto na Tabela 5.

Alguns quadros do vídeo Futebol, usado nos experimentos, foram selecionados aleatoriamente e exibidos nas Figs. 25, 26, 27 e 28 para ajudar a visualizar os resultados do ponto de vista do usuário. Com base nas análises dos gráficos, o nosso objetivo principal

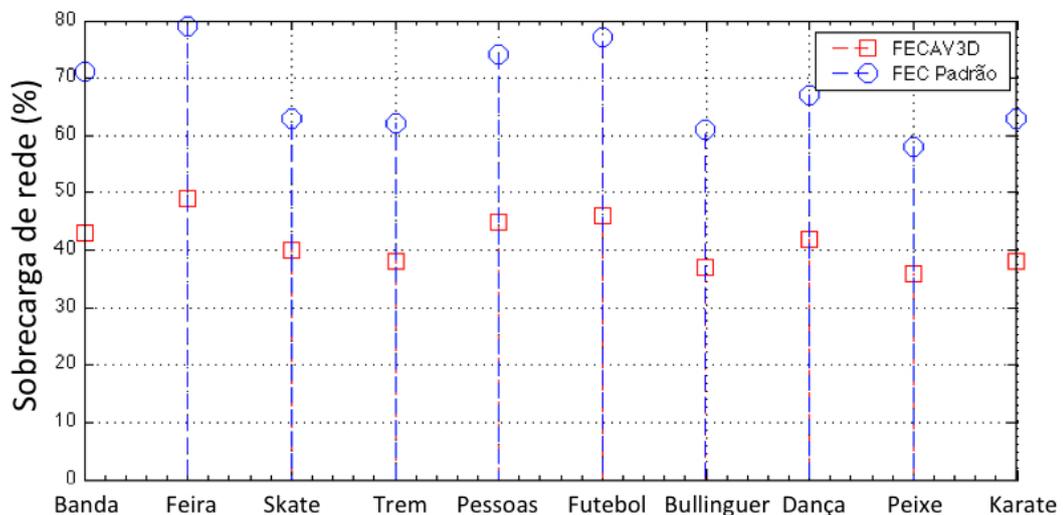


Figura 23: Sobrecarga de rede

Tabela 5: Taxa de recuperação para o vídeo 3D Futebol

Parâmetros	FEC Padrão	FEC Padrão	FECAV3D	FECAV3D
	Pior Caso	Melhor Caso	Pior Caso	Melhor Caso
Perda sem FEC (%)	20,43	23,43	21,31	23,32
Recuperação (%)	43,12	47,43	43,01	47,20
Perda com FEC (%)	11,62	10,74	11,64	10,78
SSIM	0,883	0,853	0,867	0,850
VQM	1,043	0,614	1,042	0,634
MOS	3,5	4,1	3,4	3,8

- reduzir a sobrecarga de rede e manter a boa qualidade de vídeo 3D - foi alcançado. Fomos capazes de reduzir a sobrecarga de rede em até 40,26% tendo, visualmente o vídeo 3D pelo menos a mesma, ou qualidade muito similar, como foi demonstrado pela avaliação subjetiva, que envolveu os observadores humanos.

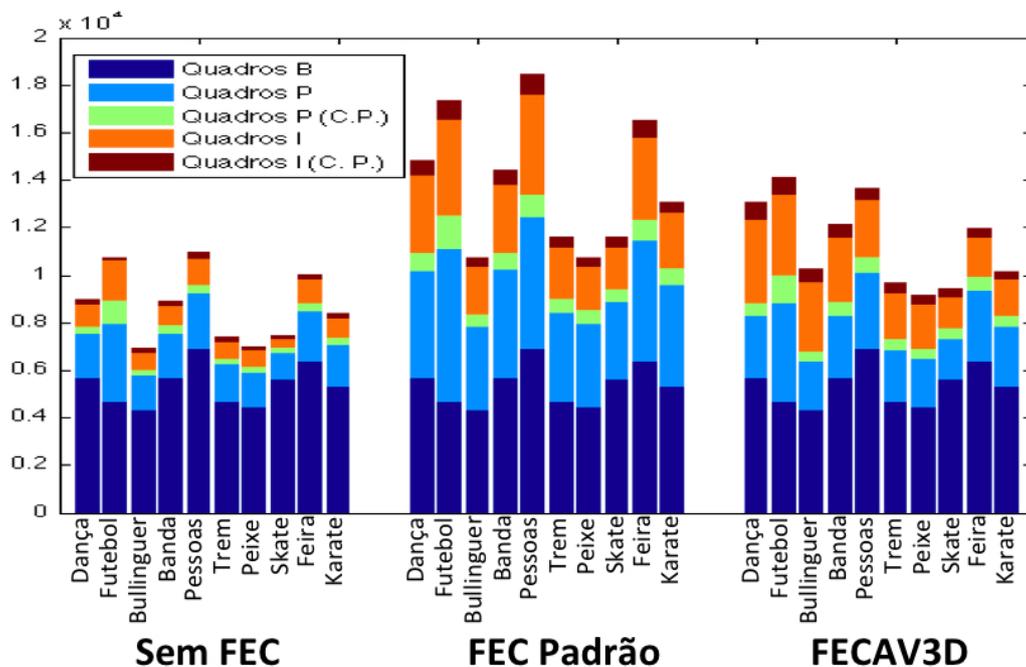


Figura 24: Sobrecarga de rede em pacotes

Figura 25: Frame 98 original com *views* esquerda e direita do vídeo Futebol

## 6.1 Conclusões

A partir da análise dos resultados obtidos, conclui-se que o mecanismo proposto consegue alcançar boas taxas de MOS, SSIM e VQM, reduzindo substancialmente a sobrecarga na rede. Com base nos resultados apresentados neste capítulo, acreditamos que o mecanismo FECAV3D mostrou bons resultados e permitiu a melhora em termos de qualidade de vídeo 3D através de WMNs. A rede em malha sem fio foi utilizada para teste e cenário das simulações. Espera-se que o mecanismo proposto consiga também, melhores resultados numa segunda fase, quando poderá ser adaptado para este tipo de rede.

O grau de codificação e atividade espacial do vídeo destacam-se como fatores



Figura 26: Frame 98 com *views* esquerda e direita do vídeo Futebol sem FEC



Figura 27: Frame 98 com *views* esquerda e direita do vídeo Futebol com FECAV3D



Figura 28: Frame 98 com *views* esquerda e direita do vídeo Futebol com FEC Padrão

significativos para o mecanismo proposto, onde, o número de pacotes do tipo B dentro de um mesmo vídeo, revela-se como parâmetro influenciador na recuperação de pacotes para o FECAV3D. Valores menores de pacotes do tipo B possibilitam uma melhor atuação do FECAV3D em função de perda de quadros P e I.

O mecanismo FECAV3D apresentou boas avaliações de QoE dos vídeos para quase todos os valores de perda, provando que técnicas para aferição e otimização de QoE em tráfegos multimídia, como o FEC podem ser utilizadas sem oneração dos recursos da rede e de tráfegos concorrentes, possibilitando uma melhora bastante desejável para as redes baseadas em IP atuais.

---

---

## CAPÍTULO 7

---

# CONCLUSÕES, TRABALHOS FUTUROS

Métodos capazes de aprimorar com sucesso o nível de QoE são vitais para o sucesso de sistemas multimídia em cenários sem fio e tem recebido muita atenção, sobretudo da indústria e da academia nos últimos anos. Além disso, novas demandas de recentes serviços tem cada vez mais crescido e exigido níveis maiores de experiência para o usuário final. Este é o caso do *3D video streaming*, *3D-on-demand* e 3DTV que são previstos a explodirem em pouco tempo como tem acontecido na indústria de dispositivos e cinema para este tipo de tecnologia.

Os atuais métodos de otimização 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 3D e 2D, 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 das avaliações dos usuários são alguns, dos muitos, fatores que podem influenciar a nota de qualidade do avaliador. Porém os mecanismos para a melhora dos níveis de QoE devem possuir tal capacidade e abrangência, uma vez que a Internet está repleta de vídeos com diferentes codificações e sujeita à diversos tipos e níveis de degradação.

Além disso, as redes em malha sem fio definem cenários que tornam 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 motivo, 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 FECAV3D ter sido implementado para uma rede em malha sem fio, o mecanismo proposto não está retrito à este cenário, uma vez que outras tecnologias de transmissão podem ser utilizadas, desde que as etapas de

implementação do FECAV3D, descritas no Capítulo 5, sejam seguidas.

Foi observado que o mecanismo FEC Adaptativo para Vídeo 3D (FECAV3D) com esquema de Proteção Desigual de Erros melhora a distribuição de vídeos 3D sem adicionar sobrecarga de rede desnecessária. Um conjunto de testes controlados foi realizado considerando sequências de vídeos 3D de diferentes tipos, graus de complexidade e movimentação. Estas características são esperadas para os vídeos na Internet. Estes vídeos foram avaliados usando métricas de QoE subjetivas e objetivas amplamente conhecidas e utilizadas pela comunidade acadêmica e pela indústria.

Os resultados das simulações mostraram que o mecanismo FECAV3D se equipara ao mecanismo FEC padrão não adaptativo em termos da taxa de recuperação e qualidade do vídeo. Em particular, no que se refere à sobrecarga de rede o FECAV3D consegue superar com grande folga o mecanismo de FEC Padrão. Através das simulações observou-se que os blocos flexíveis de FEC conseguem fazer a transmissão de vídeos 3D mais resistentes à perda de pacotes (com uma boa taxa de recuperação), e, consequentemente, obtendo uma melhor qualidade de vídeo. Além disso, observou-se que não há necessidade de proteger todos os pacotes de um GOP de forma a se obter uma melhora no nível de qualidade de vídeo, pois os codecs são resistentes a uma certa quantidade de perda, em particular, dos pacotes no final do GOP. Durante as simulações, foi perceptível uma sobrecarga na rede entre 36% e 49% (em média, 40% a menos do que a gerada pelo FEC padrão não adaptativo).

Finalmente, para trabalho futuros, abaixo estão listadas algumas tarefas a serem realizadas:

- Implementar e avaliar o FECAV3D em uma rede real e realizar experimentos em maior escala e em diferentes cenários;
- Implementar um mecanismo baseado em FEC que além de características dos vídeos, em tempo real, também faça uso de características das redes. Por exemplo, numa rede em malha sem fio, poderia se utilizar a característica multipercurso deste tipo de tecnologia.
- Adaptar o FECAV3D para aplicações como *games* que também exigem altos níveis de qualidade tanto em vídeo como em som.
- Desenvolver uma análise mais detalhada no que diz respeito à qualidade de profundidade das sequências de vídeos 3D. Não somente avaliar o impacto dos quadros I,P e B, mas também estender explicações a respeito entrega ou não da camada de profundidade e as otimizações que podem decorrer dentro da mesma para vídeos 3D.

## 7.1 Publicações

Durante o desenvolvimento da proposta desta dissertação os seguintes trabalhos foram publicados:

- QoE-based Packet Drop Control for 3D-Video Streaming - QUADROS, Carlos Jean; COSTA, A. D. B; RIKER, A; CERQUEIRA, Eduardo Coelho - Encouraging Network Research in Latin America, v. 0000, p. 0000, 2012.
- A Quality of Experience Handover System for Heterogeneous Multimedia Wireless Networks - QUADROS, Carlos Jean; PESCAPÉ, Antônio, NETO, Augusto; RIKER, A.; CURADO, Marília; CERQUEIRA, Eduardo Coelho.
- Control of QoE based on Algorithms for the Disposal of Packets concerned with Streaming Video in Wireless Networks - BEZERRA, P; MELO, A; COSTA, A; QUADROS, Carlos Jean; CERQUEIRA, Eduardo Coelho; ABELEM, A. J. - International journal of computer science and network security, v. 12, p. 1-12, 2012.
- A Mobile QoE Architecture for Heterogeneous Multimedia Wireless Networks - QUADROS, C. J. F.; RICKER, A.; IMICH, R.; NETO, A.; CURADO, M.; Cerqueira, Eduardo - International Workshop on Mobility Management in the Networks of the Future World (IEEE MobiWorld 2012), 2112.
- QoE-based Packet Dropper Controllers for Multimedia Streaming in WiMAX Networks. - COSTA, A.; QUADROS, Carlos Jean F; MELO, A; RODRIGUES, D; MONTEIRO, E; Cerqueira, Eduardo; ABELEM, A. J. - In: 6th IFIP/ACM Latin America Networking Conference, 2011, Quito. 6th IFIP/ACM Latin America Networking Conference, 2011.
- ETXMULT: a new routing metric for multimedia applications in wireless mesh networks - RIKER, A.; AGUIAR, E.; QUADROS, Carlos Jean; Cerqueira, Eduardo; ABELEM, A. J - In: IEEE 3rd Latin American Conference on Communications (IEEE LATINCOM 2011), 2011, Belém.

---

## REFERÊNCIAS

- 802.11s (2010). Ieee mesh networking. *Task Group S Status of Project IEEE 802.11s*, <http://www.ieee802.org/11/Reports/tgsupdate.htm>.
- Abelém, A., Albuquerque, C., Saade, D., Aguiar, E., Duarte, J., Da Fonseca, J., and Magalhães, L. (2007). Redes mesh: Mobilidade, qualidade de serviço e comunicação em grupo.
- Adlakha, S., Zhu, X., Girod, B., and Goldsmith, A. (2007). Joint capacity, flow and rate allocation for multiuser video streaming over wireless ad-hoc networks. In *Communications, 2007. ICC'07. IEEE International Conference on*, pages 1747–1753. IEEE.
- Aguiar, E., Riker, A., Mu, M., Zeadally, S., Cerqueira, E., and Abelem, A. (2012). Real-time qoe prediction for multimedia applications in wireless mesh networks. In *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, pages 592–596. IEEE.
- Aïache, H., Conan, V., Lebrun, L., and Rousseau, S. (2008). A load dependent metric for balancing internet traffic in wireless mesh networks. In *Mobile Ad Hoc and Sensor Systems, 2008. MASS 2008. 5th IEEE International Conference on*, pages 629–634. IEEE.
- Aksay, A., Bici, M., Bugdayci, D., Tikänmaki, A., Gotchev, A., and Akar, G. (2009). A study on the effect of mpe-fec for 3d video broadcasting over dvb-h. In *Proceedings of the 5th International ICST Mobile Multimedia Communications Conference*, page 1. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- Akyildiz, I. and Wang, X. (2005). A survey on wireless mesh networks. *Communications Magazine, IEEE*, 43(9):S23–S30.
- Akyildiz, I. and Wang, X. (2008). Cross-layer design in wireless mesh networks. *Vehicular Technology, IEEE Transactions on*, 57(2):1061–1076.
- Akyildiz, I. and Wang, X. (2009). *Wireless mesh networks*, volume 1. Wiley.

- Alay, Ö., Korakis, T., Wang, Y., and Panwar, S. (2010). Dynamic rate and fec adaptation for video multicast in multi-rate wireless networks. *Mobile Networks and Applications*, 15(3):425–434.
- Ancillotti, E., Bruno, R., Conti, M., and Pinizzotto, A. (2011). Load-aware routing in mesh networks: Models, algorithms and experimentation. *Computer Communications*, 34(8):948–961.
- 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.
- Bouhouch, H., Guemara, S., and Fatmi, E. (2007). Qos routing in ad hoc networks by integreting activity in the olsr protocol. In *Second International Conference on Systems and Networks Communications*, pages 1–1.
- Breuel, C. (2004). Redes em malha sem fio. *Instituto de Matematica e Estatistica, USP*, [http://grenoble.ime.usp.br/movel/Wireless Mesh Networks.pdf](http://grenoble.ime.usp.br/movel/Wireless%20Mesh%20Networks.pdf). Dezembro.
- Bruno, R., Conti, M., and Gregori, E. (2005). Mesh networks: commodity multihop ad hoc networks. *Communications Magazine, IEEE*, 43(3):123–131.
- Chung, Y. (2010). Symmetrical frame discard method for 3d video over ip networks. *Consumer Electronics, IEEE Transactions on*, 56(4):2790–2796.
- Couto, D., Aguayo, D., Bicket, J., and Morris, R. (2005). A high-throughput path metric for multi-hop wireless routing. *Wireless Networks*, 11(4):419–434.
- de Albuquerque, C., Saade, D., Leite, J., Magalhães, L., Teixeira, D., Neves, L., and da Fonseca, J. (2004). Gt-mesh rt1, termo de referencia e estado da arte. *Termo de referencia e estado da arte, 2006*, <http://mesh.ic.uff.br>.
- Díaz, C., Cabrera, J., Jaureguizar, F., and García, N. (2011). A video-aware fec-based unequal loss protection scheme for rtp video streaming. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pages 3–4. IEEE.
- Draves, R., Padhye, J., and Zill, B. (2004). Routing in multi-radio, multi-hop wireless mesh networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 114–128. ACM.
- Faccin, S., Wijting, C., Kenckt, J., and Damle, A. (2006). Mesh wlan networks: concept and system design. *Wireless Communications, IEEE*, 13(2):10–17.
- Fehn, C. (2004). Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv. In *Electronic Imaging 2004*, pages 93–104. International Society for Optics and Photonics.
- Fehn, C., Hopf, K., and Quante, B. (2004). Key technologies for an advanced 3d-tv system. In *Proc. of SPIE Vol*, volume 5599, page 67.
- Fehn, C., Kauff, P., De Beeck, M., Ernst, F., Ijsselsteijn, W., Pollefeys, M., Van Gool, L., Ofek, E., and Sexton, I. (2002). An evolutionary and optimised approach on 3d-tv. In *Proc. of IBC*, volume 2, pages 357–365.

- Fitzek, F. and Reisslein, M. (2001). Mpeg-4 and h. 263 video traces for network performance evaluation. *Network, IEEE*, 15(6):40–54.
- Flierl, M., Mavlankar, A., and Girod, B. (2007). Motion and disparity compensated coding for multiview video. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(11):1474–1484.
- Forecast, C. (2010). Cisco visual networking index: Global mobile data traffic forecast update 2009–2014. *Cisco Public Information, February*, 9.
- Gao, X., Lu, W., Tao, D., and Li, X. (2009). Image quality assessment based on multiscale geometric analysis. *Image Processing, IEEE Transactions on*, 18(7):1409–1423.
- Genetzakis, M. and Siris, V. (2008). A contention-aware routing metric for multi-rate multi-radio mesh networks. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2008. SECON'08. 5th Annual IEEE Communications Society Conference on*, pages 242–250. IEEE.
- Greengrass, J., Evans, J., and Begen, A. (2009). Not all packets are equal, part i: Streaming video coding and sla requirements. *Internet Computing, IEEE*, 13(1):70–75.
- Gualdi, G., Prati, A., and Cucchiara, R. (2008). Video streaming for mobile video surveillance. *Multimedia, IEEE Transactions on*, 10(6):1142–1154.
- 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.
- Han, L., Park, S., Kang, S., and In, H. (2010). An adaptive fec mechanism using crosslayer approach to enhance quality of video transmission over 802.11 w lans. *KSII Trans. Internet Inform. Syst*, 4(3):341–357.
- Hanzo, L., Cherriman, P., and Streit, J. (2007). *Video compression and communications: from basics to H. 261, H. 263, H. 264, MPEG4 for DVB and HSDPA-style adaptive turbo-transceivers*. Wiley-IEEE Press.
- HARADA, E. (2006). Wireless mesh networks: Uma tecnologia que promete. <http://sisnema.com.br/Materias/idmat017459.htm>.
- Haskell, B., Puri, A., and Netravali, A. (1996). *Digital video: an introduction to MPEG-2*. Springer.
- Hassan, M. and Landolsi, T. (2010). A retransmission-based scheme for video streaming over wireless channels. *Wireless Communications and Mobile Computing*, 10(4):511–521.
- Hewage, C., Nasir, S., Worrall, S., and Martini, M. (2010). Prioritized 3d video distribution over ieee 802.11 e. In *Future Network and Mobile Summit, 2010*, pages 1–9. IEEE.
- Index, C. (2012). Forecast and methodology, 2011-2016.
- ITU-R-B500-11 (2002). *Methodology for the Subjective Assessment of the Quality of Television Pictures*. International Telecommunication Union, Geneva, Switzerland.

- Jacquet, P. (2003). Optimized link state routing protocol (olsr).
- Jean, C., Melo, A., Douglas, A., Abelém, A., Cerqueira, E., Neto, A., Riker, A., and Curado, M. (2012). Qoe-based packet drop control for 3d-video streaming over wireless networks. In *Proceedings of the 7th Latin American Networking Conference*, pages 59–66. ACM.
- Kalva, H., Christodoulou, L., Mayron, L., Marques, O., and Furht, B. (2006). Challenges and opportunities in video coding for 3d tv. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 1689–1692. IEEE.
- Kaup, A. and Fecker, U. (2006). Analysis of multi-reference block matching for multi-view video coding. In *Proc. 7th Workshop Digital Broadcasting*, pages 33–39. Citeseer.
- Khan, A., Sun, L., and Ifeachor, E. (2009). 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.
- Klaue, J., Rathke, B., and Wolisz, A. (2003). Evalvid—a framework fo video transmission and quality evaluation. *Computer Performance*, pages 255–272.
- Koumaras, H., Lin, C., Shieh, C., 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.
- Le, T., Nguyen, H., and Zhang, H. (2010). Evalsvc, an evaluation platform for scalable video coding transmission. In *Consumer Electronics (ISCE), 2010 IEEE 14th International Symposium on*, pages 1–6. IEEE.
- Lee, J., Chen, C., Horng, M., and Kuo, Y. (2011). An efficient adaptive fec algorithm for short-term quality control in wireless networks. In *Advanced Communication Technology (ICACT), 2011 13th International Conference on*, pages 1124–1129. IEEE.
- Lindeberg, M., Kristiansen, S., Plagemann, T., and Goebel, V. (2011). Challenges and techniques for video streaming over mobile ad hoc networks. *Multimedia Systems*, 17(1):51–82.
- Liu, T. and Liao, W. (2009). Interference-aware qos routing for multi-rate multi-radio multi-channel ieee 802.11 wireless mesh networks. *Wireless Communications, IEEE Transactions on*, 8(1):166–175.
- Lo, C., Lin, C., Chen, Y., and Yu, J. (2011). Contribution-based peer selection for packet protection for p2p video streaming over mesh-based networks. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 2233–2236. IEEE.
- Luiz Jr, O. and Przybysz, A. (2007). Infra-estrutura e roteamento em redes wireless mesh. *Synergismus scyentifica UTFPR*, 2(1).
- Ma, L. and Denko, M. (2007). A routing metric for load-balancing in wireless mesh networks. In *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*, volume 2, pages 409–414. IEEE.

- Merkle, P., Müller, K., Smolic, A., and Wiegand, T. (2005). Statistical evaluation of spatiotemporal prediction for multi-view video coding. *Proc. ICOB*, 2005:27–28.
- Mueller, K., Merkle, P., Schwarz, H., Hinz, T., Smolic, A., Oelbaum, T., and Wiegand, T. (2006). Multi-view video coding based on h. 264/avc using hierarchical b-frames. In *Picture Coding Symposium*, volume 2006.
- Nafaa, A., Taleb, T., and Murphy, L. (2008). Forward error correction strategies for media streaming over wireless networks. *Communications Magazine, IEEE*, 46(1):72–79.
- Neckebroek, J., Moeneclaey, M., and Magli, E. (2010). Comparison of reed-solomon and raptor codes for the protection of video on-demand on the erasure channel. In *Information Theory and its Applications (ISITA), 2010 International Symposium on*, pages 856–860. IEEE.
- Oh, B. and Chen, C. (2010). A cross-layer adaptation hcca mac for qos-aware h. 264 video communications over wireless mesh networks. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 2259–2262. IEEE.
- Oh, K. and Ho, Y. (2006). Multi-view video coding based on the lattice-like pyramid gop structure. In *Proc. PCS*.
- P.900, I.-T. (2002). *Subjective video quality assessment methods for multimedia applications*. International Telecommunication Union, Geneva, Switzerland.
- Pal, A. and Nasipuri, A. (2011). A quality based routing protocol for wireless mesh networks. *Pervasive and Mobile Computing*, 7(5):611–626.
- 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.
- Parissidis, G., Karaliopoulos, M., Spyropoulos, T., and Plattner, B. (2011). Interference-aware routing in wireless multihop networks. *Mobile Computing, IEEE Transactions on*, 10(5):716–733.
- Piamrat, K., Viho, C., Bonnin, J., and Ksentini, A. (2009). 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. IEEE.
- Pinson, M. H. and Wolf, S. (2004). A new standardized method for objectively measuring video quality. *Broadcasting, IEEE Transactions on*, 50(3):312–322.
- Rec, I. (2007). P. 10/g. 100 amendment 1, new appendix i, definition of quality of experience (qoe). *International Telecommunication Union*.
- Salyers, D., Striegel, A., and Poellabauer, C. (2008). Wireless reliability: Rethinking 802.11 packet loss. In *World of Wireless, Mobile and Multimedia Networks, 2008. WoWMoM 2008. 2008 International Symposium on a*, pages 1–4. IEEE.

- Schwarz, H., Marpe, D., and Wiegand, T. (2006). Analysis of hierarchical b pictures and mctf. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 1929–1932. IEEE.
- Seshadrinathan, K., Soundararajan, R., Bovik, A., and Cormack, L. (2010). Study of subjective and objective quality assessment of video. *Image Processing, IEEE Transactions on*, 19(6):1427–1441.
- Shum, H., Kang, S., and Chan, S. (2003). Survey of image-based representations and compression techniques. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(11):1020–1037.
- Smolic, A., Mueller, K., Stefanoski, N., Ostermann, J., Gotchev, A., Akar, G., Triantafyllidis, G., and Koz, A. (2007). Coding algorithms for 3d tv a survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(11):1606–1621.
- Song, Z., Wang, H., Wen, Y., Wu, D., and Lee, K. (2012). Depth-color based 3d image transmission over wireless networks with qoe provisions. *Computer Communications*.
- Staelens, N., Sedano, I., Barkowsky, M., Janowski, L., Brunnstrom, K., and Le Callet, P. (2011). Standardized toolchain and model development for video quality assessment, the mission of the joint effort group in vqeg. In *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, pages 61–66. IEEE.
- Takahashi, A., Hands, D., and Barriac, V. (2008). Standardization activities in the itu for a qoe assessment of iptv. *Communications Magazine, IEEE*, 46(2):78–84.
- Tsai, M., Chilamkurti, N., Zeadally, S., and Vinel, A. (2011). Concurrent multipath transmission combining forward error correction and path interleaving for video streaming. *Computer Communications*, 34(9):1125–1136.
- Tsirigos, A. and Haas, Z. (2004). Analysis of multipath routing-part i: The effect on the packet delivery ratio. *Wireless Communications, IEEE Transactions on*, 3(1):138–146.
- Uemura, S., Fukumoto, N., Yamada, H., and Nakamura, H. (2008). Qos/qoe measurement system implemented on cellular phone for ngn. In *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*, pages 117–121. IEEE.
- Vatolin, D., Moskvina, A., Petrov, O., and Trunichkin, N. (2009). Msu video quality measurement tool.
- Vercammen, N., Staelens, N., Rombaut, A., Vermeulen, B., and Demeester, P. (2008). Extensive video quality evaluation: a scalable video testing platform. In *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*, pages 91–97. IEEE.
- Waharte, S., Boutaba, R., Iraqi, Y., and Ishibashi, B. (2006). Routing protocols in wireless mesh networks: challenges and design considerations. *Multimedia Tools and Applications*, 29(3):285–303.
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004a). Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612.

- Wang, Z., Bovik, A. C., and Sheikh, H. R. and Simoncelli, E. P. (2004b). Image quality assessment: From error measurement to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612.
- Watkinson, J. (2012). *The MPEG handbook*. Focal Press.
- Weingartner, E., Vom Lehn, H., and Wehrle, K. (2009). A performance comparison of recent network simulators. In *Communications, 2009. ICC'09. IEEE International Conference on*, pages 1–5. IEEE.
- Wilhelmsson, L. and Milstein, L. (1999). On the effect of imperfect interleaving for the gilbert-elliott channel. *Communications, IEEE Transactions on*, 47(5):681–688.
- Wilibox (2009). Wireless linux in the box. *Task Group S. Status of Project IEEE 802.11s*, <http://www.wilibox.com/products/wilimesh>.
- Winkler, S. (2009). Video quality measurement standards, current status and trends. In *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, pages 1–5. IEEE.
- 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.
- Yang, Y., Wang, J., and Kravets, R. (2006). Load-balanced routing for mesh networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 10(4):3–5.
- Yuan, Y., Cockburn, B., Sikora, T., and Mandal, M. (2006). A gop based fec technique for packet based video streaming. In *Proceedings of the 10th WSEAS international conference on Communications*, pages 187–192. World Scientific and Engineering Academy and Society (WSEAS).
- 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.
- Zhu, R. (2011). Intelligent rate control for supporting real-time traffic in wlan mesh networks. *Journal of Network and Computer Applications*, 34(5):1449–1458.
- Zwinkels, P. (2006). Wireless mesh network. *Achieving the Triple Play: Technologies and Business Models for Success*, page 305.