

Rogério Pernas Nunes

**CODIFICAÇÃO ESCALÁVEL DE VÍDEO
PARA RECEPÇÃO FIXA NO SISTEMA
BRASILEIRO DE TELEVISÃO DIGITAL**

Dissertação apresentada à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Mestre em Engenharia.

São Paulo
2009

Rogério Pernas Nunes

**CODIFICAÇÃO ESCALÁVEL DE VÍDEO
PARA RECEPÇÃO FIXA NO SISTEMA
BRASILEIRO DE TELEVISÃO DIGITAL**

Dissertação apresentada à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Mestre em Engenharia.

Área de concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. Marcelo Knörich Zuffo

São Paulo
2009

Dedicatória

A meus pais, Maria Aparecida P. Nunes e Sérgio Nunes, com muito amor e gratidão, pelos verdadeiros amigos que são em todos os momentos de minha vida.

Agradecimentos

À minha família e a Mariana Stefanini, pelo fundamental apoio ao longo deste trabalho, especialmente nos momentos mais difíceis.

Ao Prof. Dr. Marcelo Knörich Zuffo, pela orientação e pela confiança contínua no meu trabalho.

Ao Sr. Hilel Becher, gerente do Núcleo de Engenharia de Mídias do Laboratório de Sistemas Integráveis, que foi muito compreensivo, disponibilizando apoio e recursos para a realização deste trabalho.

Aos colegas do Laboratório de Sistemas Integráveis que me ajudaram com a experiência de seus Mestrados, Célio Hira, Eduardo Carvalho, Gil Barros.

Aos demais colegas de trabalho do Laboratório de Sistemas Integráveis que me ajudaram de diversas formas, com apoio técnico e amizade. Em especial a Rafael Herrero, que me ajudou com a camada de transporte do SBTVD.

A Heiko Schwarz (Fraunhofer HHI) e Vincent Bottreau (Thomson) que foram bastante solícitos no esclarecimento de dúvidas sobre o SVC.

Ao Prof. Dr. Guido Stolfi, que além de ter lecionado a disciplina de TV Digital durante minha graduação, trouxe esclarecimentos e contribuições a este trabalho.

Financiamento

Este trabalho contou com apoio financeiro da Associação do Laboratório de Sistemas Integráveis Tecnológico (LSI-TEC) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por meio da Rede de Cooperação Universitária para Ensino Superior e Pesquisa Avançada em TV Digital e seus Cenários Interdisciplinares de Evolução.

Resumo

Em dezembro de 2007, a partir da cidade de São Paulo, as transmissões de televisão digital terrestre e aberta tiveram início no Brasil. Um avanço significativo do Sistema Brasileiro de TV Digital (SBTVD) foi a adoção do padrão H.264/AVC e o formato de vídeo 1080i para a codificação de vídeo em alta definição. A adoção em larga escala de tecnologia de alta definição tem sido um processo observado em vários mercados do mundo, e novos formatos superiores ao 1080i já estão sendo discutidos e propostos. Tendo em vista o que será a próxima geração da televisão, centros de pesquisa, como o centro japonês da emissora NHK, investigam os fatores humanos determinantes para caracterizar o sistema que deverá ser o último passo em tecnologia de televisão 2D. Já nomeado de UHDTV, este sistema deve contemplar resolução de 7680 pontos horizontais por 4320 pontos verticais, além de outras características ainda em estudo. Ao mesmo tempo, o trabalho aqui apresentado discute as ferramentas de suporte à escalabilidade na codificação multimídia como forma de evolução gradual dos formatos de vídeo na radiodifusão. Especificamente este trabalho sistematiza as ferramentas de escalabilidade do padrão H.264/AVC tendo em vista a sua aplicação ao SBTVD. Neste sentido, são discutidas as possibilidades de evolução do sistema frente à escalabilidade e são apresentados levantamentos experimentais da atual ocupação do espectro na cidade de São Paulo, evidenciando a disponibilidade de taxa para expansões. São apresentados também resultados iniciais relativos à codificação SVC, que apontam objetivamente as vantagens da escalabilidade sobre o *simulcast*, evidenciando que esta técnica pode ser utilizada no SBTVD para prover novos formatos de vídeo, tendo como premissa a compatibilidade com os atuais receptores que suportam o formato 1080i. O trabalho apresenta contribuições teóricas e experimentais na direção de adoção da escalabilidade no SBTVD, apontando também possíveis trabalhos futuros que, se realizados, poderão confirmar a transmissão de formatos superiores de vídeo nos próximos anos no SBTVD.

Palavras-chave: Codificação escalável de vídeo. SBTVD. TV digital.

Abstract

On December 2007, starting from São Paulo city, the open digital terrestrial transmissions were launched in Brazil. A significant improvement of the Brazilian Digital TV System (SBTV D) was the adoption of the H.264/AVC standard supporting the 1080i video format for the high definition video coding. Wide adoption of high definition technology has been a process that can be observed in lots of countries, and new video formats, beyond 1080i have already been discussed and proposed. With both eyes in the next generation of TV, research centers like Japanese broadcaster NHK investigate human factors that should drive the system specifications of this one that may be the last step in terms of 2D television technology. Named UHDTV, the system may support 7680 horizontal dots per 4320 vertical dots in terms of resolution among other features. At the same time, the work exposed here discusses tools that support multimedia coding scalability as a way of gradually improving video formats for broadcast. This work specially deals with the H.264/AVC standard scalability tools, aiming their use within SBTVD. Therefore, the evolution of the system is discussed based on scalability and experimental results related to the digital TV spectral occupation in São Paulo city are analyzed, showing that there is enough exceeding bit rate available for future expansion. Initial results related to SVC coding are also shown, objectively indicating that video scalability is more advantageous than simulcast and that this technique can be used in SBTVD to provide new video formats, keeping compatibility with current receivers that only support 1080i format. This work presents theoretical and experimental contributions towards the adoption of SVC in the SBTVD, pointing out some future works that, if executed, could confirm the transmission of new video format in SBTVD in the next years.

Keywords: Scalable video coding. SBTVD. Digital TV.

Lista de Figuras

1.1	Aparato de transmissão do primeiro sistema televisivo de John Baird	23
1.2	Representação parcial do sistema Super Hi-Vision	27
2.1	Principais técnicas de compressão digital dos anos 50 aos 80	37
2.2	Normas de codificação de vídeo publicadas pela ITU e pela ISO/IEC	39
2.3	Estrutura genérica de um codificador híbrido	40
2.4	Marcos da codificação de vídeo	41
2.5	Esquema de codificação escalável por camadas	44
2.6	Arquitetura do codificador de vídeo H.264/AVC	50
2.7	Tamanhos de blocos de predição de movimento no H.264/AVC	51
2.8	Modos de predição intra de macroblocos subdivididos em blocos 4x4 na norma H.264/AVC	52
2.9	Escalabilidade temporal dicotômica no SVC	61
2.10	Escalabilidade temporal não dicotômica no SVC	61
2.11	Escalabilidade temporal com atraso nulo no SVC	61
2.12	Escalabilidade espacial no SVC com predição entre camadas	63
2.13	Tipos de estruturas de referência na escalabilidade por qualidade do SVC .	65
3.1	Processo de definição da TV digital no Brasil	72
3.2	Cronograma de implantação da TV digital terrestre no Brasil	76
3.3	Normas que compõem o SBTVD-T	76
3.4	Exemplo de uso do canal de 6 MHz no SBTVD-T e bandas de guarda	78
3.5	Diagrama de blocos de transmissão do SBTVD-T	80
3.6	Espectro alocado pela ANATEL para o SBTVD	80

3.7	Relação entre sinais fonte, pacotes PES, tabelas SI e fluxo TS	85
3.8	Estrutura de um pacote PES	85
3.9	Estrutura de pacotes de um TS	86
4.1	Tela de análise de um canal COFDM do software da Fujitsu	104
4.2	Telas de análise com software MTS400 da Tektronix	105
4.3	Seqüências de vídeo para testes	113
4.4	Etapas de geração das seqüências de vídeo	114
5.1	Exemplos de imagens obtidas nos processos de subamostragem para geração das seqüências de vídeo	121
5.2	Comparação de qualidade entre diferentes estruturas de codificação (Formato 720p)	132
5.3	Comparação de qualidade entre diferentes estruturas de codificação (Formato 1080p)	133
5.4	Resultados da codificação por camada simples (720p, I=48, GOP=8) . . .	134
5.5	Resultados da codificação por camada simples (1080p, I=48, GOP=8) . . .	135
5.6	Resultados da codificação por camada simples (720p, I=24, GOP=8) . . .	136
5.7	Resultados da codificação por camada simples (1080p, I=24, GOP=8) . . .	137
5.8	Resultados da codificação por camada simples (720p, I=24, GOP=4) . . .	138
5.9	Resultados da codificação por camada simples (1080p, I=24, GOP=4) . . .	139
A.1	Sistemática de captura de vídeo analógico	148
A.2	Entrelaçamento do sinal de vídeo	150

Lista de Tabelas

2.1	Efeitos comuns em compressão e codificação de vídeo	36
2.2	Ferramentas dos perfis básicos e da extensão FRExt do H264/AVC	57
2.3	Níveis do H.264/AVC	58
3.1	Lista de normas do SBTVD-T	77
3.2	Taxas máxima e mínima por serviço do canal de exemplo	79
3.3	Resoluções de luminância com suporte obrigatório para SDTV e HDTV	82
3.4	Taxas de quadros com suporte obrigatório para SDTV e HDTV	82
3.5	Parâmetros de colorimetria recomendados para SDTV e HDTV	83
3.6	Resoluções de luminância com suporte obrigatório para LDTV	83
3.7	Taxas de quadros com suporte obrigatório para LDTV	83
3.8	Tabelas principais do TS no SBTVD-T	86
4.1	Resultados dos testes de escalabilidade do JVT para HDTV	91
4.2	Dificuldade de codificação das seqüências de vídeo	112
5.1	Configuração de transmissão para receptores fixos em São Paulo	117
5.2	Características dos fluxos principais de vídeo das emissoras de São Paulo	119
5.3	PSNR das seqüências de teste para as resoluções 720p, 1080i, 1080p.	121
5.4	Qualidade PSNR média das seqüências 720p/1080p com o codificador AteMe124	124
5.5	Coefficientes derivados pelo JSVM para as camadas temporais	126
5.6	Níveis de codificação alocados pelo JSVM	126
5.7	Resultados da codificação da seqüência CrowdRun em camada simples com o JSVM	128
5.8	Referências das seqüências em 1080p para a codificação escalável	129

5.9	Resultados do <i>simulcast</i> de 720p e 1080p	130
5.10	Resultados da escalabilidade de 720p para 1080p	130

Lista de Abreviaturas

- ABNT** Associação Brasileira de Normas Técnicas
- ANATEL** Agência Nacional de Telecomunicações
- ASO** Ordenamento Arbitrário de Fatias (*Arbitrary Slice Ordering*)
- ATSC** *Advanced Television Systems Committee*
- AVI** *Audio Video Interleave*
- BST-OFDM** *Band Segmented Transmission Orthogonal Frequency Division Multiplexing*
- CABAC** Codificação Binária Aritmética Adaptada ao Contexto (*Context-Adaptive Binary Arithmetic Coding*)
- CAVLC** Codificação por Comprimento Variado Adaptada ao Contexto (*Context-Adaptive Variable Length Code*)
- CGS** Escalabilidade com Baixa Granularidade (*Coarse-Grain Quality Scalability*)
- CODEC** *CO*odificador *DEC*odificador
- CPqD** Centro de Pesquisa e Desenvolvimento
- CVS** Sistema de Versões Concorrentes (*Concurrent Version System*)
- DCT** Transformada Discreta do Cosseno (*Discrete Cosine Transform*)
- DPCM** Modulação Diferencial por Código de Pulsos (*Differential Pulse Code Modulation*)
- DVB** *Digital Video Broadcasting*
- DVB-SPI** *DVB Synchronous Parallel Interface*
- EBU** *European Broadcasting Union*
- FGS** Escalabilidade por Granularidade Fina (*Fine Grain Scalability*)

ES Fluxo Elementar (*Elementary Stream*)

FINEP Financiadora de Estudos e Projetos

FMO Ordenamento Flexível de Macroblocos (*Flexible Macroblock Ordering*)

FRExt Extensão de Faixa de Fidelidade (*Fidelity Range Extensions*)

FUNTTTEL Fundo Tecnológico para o Desenvolvimento das Telecomunicações

GOP Grupo de Figuras (*Group of Pictures*)

H.264/AVC *Advanced Video Coding*. Padrão de compressão com perdas de vídeo especificado tanto pela ITU, onde é referenciado por ITU-T H.264, quanto na ISO/IEC, pertencendo à família do MPEG-4, onde é referenciado como ISO/IEC 14496-10.

HDMI *High-Definition Multimedia Interface*

HDTV Televisão de Alta Definição (*High Definition Television*)

HE-AAC v2 *High-Efficiency Advanced Audio Coding version 2*. Padrão de compressão com perdas de áudio especificado pela ISO/IEC como parte da família MPEG-4. Sua referência é ISO/IEC 14496-3.

HVS Sistema Visual Humano (*Human Visual System*)

iDCT Transformação Inversa do Cosseno (*Inverse Discrete Cosine Transform*)

IDR *Instantaneous Decoding Refresh*

ISDB *Integrated Services Digital Broadcasting*

ISDB-T *Integrated Services for Broadcasting - Terrestrial*

ISDB-TB *Integrated Services for Broadcasting - Terrestrial Brazil*

JSVM *Joint Scalable Video Model*

JVT *Joint Video Team*, formado por especialistas da ITU e da ISO/IEC

LDTV Televisão de Baixa Definição (*Low Definition Television*)

LSI Laboratório de Sistemas Integráveis

MBAFF *Macroblock Adaptive Frame Field*

MCTF Filtragem Temporal com Compensação de Movimento (*Motion-Compensated Temporal Filtering*)

MGS Escalabilidade com Granularidade Média (*Medium-Grain Quality Scalability*)

MOS Índice de Opinião Média (*Mean Opinion Score*)

MPEG *Moving Picture Experts Group*

MSE Erro Quadrático Médio (*Mean Squared Error*)

NAL Camada de Abstração de Rede (*Network Abstraction Layer*)

OFDM *Orthogonal Frequency Division Multiplexing*

PAFF *Picture Adaptive Frame Field*

PAL Linhas com Alternância de Fase (*Phase Alternation Line-rate*)

PCM Modulação por Código de Pulsos (*Pulse Code Modulation*)

PES *Packetized Elementary Stream*

PID Identificador de Pacote (*Packet Identifier*)

PS Fluxo de Programa (*Program Stream*)

PSI Informação Específica de Programa (*Program Specific Information*)

PSNR Razão Pico Sinal-Ruído (*Peak Signal-to-Noise Ratio*)

RAP Ponto de Acesso Aleatório (*Random Access Point*)

ROI Região de Interesse (*Region of Interest*)

RS Fatias Redundantes (*Redundant Slices*)

SBTVD Sistema Brasileiro de TV Digital

SBTVD-T Sistema Brasileiro de TV Digital Terrestre

SDTV Televisão de Definição Padrão (*Standard Definition Television*)

SEI Informações Suplementares de Enriquecimento (*Supplemental Enhancement Information*)

SI Informação de Sistema (*System Information*)

SNR Razão Sinal-Ruído (*Signal-to-Noise Ratio*)

SoC *System on Chip*

SSIM Similaridade Estrutural (*Structural SIMilarity*)

SVC *Scalable Video Coding*, extensão de codificação escalável de vídeo do H.264/AVC

SVT *Sveriges Television*

TAR Terminal de Acesso de Referência

TS Fluxo de Transporte (*Transport Stream*)

UHDTV Sistema de TV de Ultra Alta Definição (*Ultra High Definition TV*), conhecido também por *Super Hi-Vision*

UHF *Ultra High Frequency*, faixa de frequência do espectro de telecomunicações que vai de 300 MHz à 3 GHz

VCEG *Video Coding Experts Group*

VCL Camada de Codificação de Vídeo (*Video Coding Layer*)

VHF *Very High Frequency*, faixa de frequência do espectro de telecomunicações que vai de 30 à 300 MHz

VLC Código de Comprimento Variado (*Variable Length Code*)

VLD Decodificador de Comprimento Variado (*Variable Length Decoder*)

VQEG *Video Quality Experts Group*

VUI Informações de Usabilidade de Vídeo (*Video Usability Information*)

Sumário

1	Introdução	21
1.1	Objetivos	22
1.2	Justificativa Histórica	23
1.3	Motivação	27
1.4	Relevância	29
1.5	Estrutura da Dissertação	29
2	Codificação Escalável de Vídeo	31
2.1	Compressão e Codificação Digital	32
2.2	Análise de Qualidade de CODECs de Vídeo	33
2.3	Evolução dos CODECs de vídeo	36
2.3.1	Codificadores Híbridos ISO/ITU	39
2.3.2	Codificadores Baseados em <i>Wavelets</i>	42
2.3.3	Alguns Trabalhos Atuais	42
2.4	Escalabilidade de Vídeo	43
2.4.1	Escalabilidade Espacial	46
2.4.2	Escalabilidade Temporal	47
2.4.3	Escalabilidade por Qualidade	47
2.4.4	Escalabilidade por Controle Granular de Taxa	48
2.5	Norma de Codificação H.264/AVC	48
2.5.1	Principais Características e Funcionalidades	49
2.5.2	Perfis (<i>Profiles</i>)	55

2.5.3	Níveis (<i>Levels</i>)	58
2.6	A Extensão de Escalabilidade no H.264/AVC	59
2.6.1	Escalabilidade Temporal	60
2.6.2	Escalabilidade Espacial (ou Escalabilidade de Resolução)	62
2.6.3	Escalabilidade por Qualidade	64
2.6.4	Aplicação do SVC	67
2.6.5	Perfis e Níveis do <i>Scalable Video Coding</i> (SVC)	68
2.7	Conclusão	68
3	Sistema Brasileiro de TV Digital	70
3.1	TV no Brasil: das Transmissões Analógicas ao SBTVD	71
3.1.1	Primeiros Trabalhos e Testes com Sistemas Digitais de Vídeo	71
3.1.2	SBTVD e a Decisão do Sistema	73
3.1.3	Fórum SBTVD e Implantação do Sistema	74
3.2	Visão Geral das Normas do SBTVD-T	76
3.2.1	Sistema de Transmissão (N01)	78
3.2.2	Codificação de Vídeo (N02 Parte 1)	80
3.2.3	Sistemas de Multiplexação de Sinais (N02 Parte 3)	84
3.3	Conclusão	87
4	Considerações sobre a Escalabilidade de Vídeo para o Sistema Brasileiro de TV Digital	89
4.1	Trabalhos Correlatos	90
4.1.1	Resultados de Testes de Escalabilidade	90
4.1.2	Estágio Recente do Sistema Super Hi-Vision	92
4.1.3	Suporte à Escalabilidade no Sistema Europeu	94
4.2	Considerações sobre o Sistema Brasileiro	96
4.2.1	Receptores	98

4.2.2	Particularidades dos Formatos de Vídeo	98
4.2.3	Possibilidades com a Escalabilidade	100
4.3	Ambiente de Investigação Experimental	102
4.3.1	Requisitos	102
4.3.2	Sintonizadores Digitais	103
4.3.3	Analisadores de Fluxos de Transporte e Vídeo	104
4.3.4	Software de Referência do SVC	106
4.3.5	Outros Equipamentos e Softwares	109
4.3.6	Seqüências de Vídeo	111
4.4	Conclusão	114
5	Levantamentos Experimentais	116
5.1	Ocupação do Espectro da Cidade de São Paulo	116
5.2	Qualidade de Codificação	120
5.2.1	Métodos de Subamostragem	120
5.2.2	Estruturas Tradicionais e Estrutura B-Hierárquica	122
5.2.3	Estruturas B-Hierárquicas	124
5.3	Comparação Entre Escalabilidade e <i>Simulcast</i>	128
5.4	Transmissão para Receptor Legado	130
5.5	Conclusão	131
6	Conclusões	140
6.1	Contribuições	141
6.2	Trabalhos Futuros	142
	Referências Bibliográficas	144

Apêndices	148
A Representação de Vídeo	148
A.1 Entrelaçamento	149
A.2 Composição do Sinal de Vídeo	151
A.3 Representações Digitais	152
B Códigos-Fontes	154
B.1 Roteiro de Configuração de Variáveis	154
B.2 Roteiros de Preparação das Sequências de Vídeo	155
B.2.1 Cópia dos Arquivos Originais do Servidor	155
B.2.2 Conversão de Formato Sgi16 em I420	156
B.2.3 Truncamento das Sequências	156
B.2.4 Encapsulamento de Sequências Truncadas em Camada de Transporte AVI	158
B.3 Visualização das Sequências I420	159
B.4 Comparação entre Estruturas Tradicionais e Estrutura HBF	159
B.4.1 Configuração do Codificador AteMe	159
B.4.2 Geração dos Arquivos de Configuração para o Codificador AteMe	161
B.4.3 Geração das Tabelas de PSNR	162
B.5 Comparação Entre Estruturas HBF em Camada Simples	163
B.5.1 Configuração do JSVM para Camada Simples em 720p	163
B.5.2 Configuração do JSVM para Camada Simples em 1080p	164
B.5.3 Roteiros para Codificação das Camadas Simples com o JSVM	165
B.5.4 Configuração do AVISynth para Medição SSIM	167
B.5.5 Roteiro para Geração das Medidas SSIM das Codificações Simples com o JSVM	167
B.6 Comparação Entre Escalabilidade e <i>Simulcast</i>	168

B.6.1	Configuração do JSVM para Codificação Escalável	168
B.6.2	Roteiro para Codificação Escalável com o JSVM	169

1 Introdução

Em dezembro de 2007, as transmissões terrestres de televisão digital aberta tiveram início oficial no Brasil. A partir da cidade de São Paulo, estas transmissões iniciaram efetivamente a modernização do meio de comunicação social mais disseminado do país (CONSELHO NACIONAL DE ENGENHARIA, 2007). Com o sistema digital, tornam-se possíveis novos formatos de conteúdo, novos serviços e ainda melhor aproveitamento do espectro de radiofrequências devido às técnicas de processamento de sinal que podem ser empregadas.

Neste âmbito, o SBTVD, como foi denominado o sistema especificado e em implantação no Brasil, definiu o uso de soluções no estado-da-arte, em particular, para o processamento dos sinais de áudio (padrão HE-AAC v2) (ABNT, 2007c) e vídeo (padrão H.264/AVC) (ABNT, 2007b). Para a época, essa foi uma definição arrojada, pois nenhum outro sistema de radiodifusão aberta no mundo havia implantado com sucesso essas soluções específicas até então. Apesar disso, como a evolução tecnológica ocorre em ritmo acelerado, desde o lançamento da TV Digital surgiram diversas evoluções tecnológicas que poderiam tornar o sistema ainda melhor.

O processo de migração do sistema analógico para o digital, no entanto, é lento, pois trata-se de uma ruptura tecnológica. O novo sistema de transmissão, de forma geral, não é compatível com os equipamentos do sistema anterior, o que obriga toda a cadeia de serviço de televisão, dos radiodifusores aos usuários finais, a atualizar equipamentos.

Dessa forma, ainda que sejam desenvolvidas melhorias tecnológicas para o sistema de radiodifusão digital, se estas não forem incrementais, mas sim disruptivas como a transição analógico-digital, terão pouco apelo de implantação pelo menos ao longo dos próximos anos ou décadas, até que os investimentos do sistema atual tenham sido amortizados.

Neste contexto, este trabalho discute e avalia a escalabilidade de vídeo do H.264/AVC (ITU-T, 2007), recentemente padronizada, como solução incremental para evolução do SBTVD, possibilitando a transmissão opcional de formatos de vídeo superiores aos suportados atualmente.

Escalabilidade de vídeo, de forma geral, refere-se à capacidade de extração de partes do fluxo de bits de um vídeo codificado para adaptá-lo às diferentes necessidades e/ou preferências do usuário final, assim como também às diferentes capacidades de receptores e/ou às condições das redes de comunicação.

Com a possibilidade de aplicação desta solução à radiodifusão digital, os radiodifusores e usuários não são obrigados a trocar equipamentos em um período pré-determinado, e o processo de migração pode ser melhor controlado pela disponibilidade tecnológica e pelos interesses do mercado. Quando um radiodifusor optar por iniciar transmissões com novos padrões que contemplem escalabilidade de vídeo, espera-se que seu sinal continue compatível com os receptores em uso, fornecendo conteúdo diferenciado para usuários que optarem por atualizar seus receptores, sem deixar de atender aos demais usuários.

1.1 **Objetivos**

O objetivo deste trabalho é pesquisar e desenvolver contribuições relativas à análise e viabilidade de transmissão de vídeo escalável para receptores fixos no SBTVD. Atualmente, o formato de maior resolução especificado para esse tipo de receptor é o 1080i¹ e a escalabilidade de vídeo pode ser fundamental como solução para transmissão de formatos com resolução superior.

Dentre as diversas questões que serão abordadas, é determinante considerar a compatibilidade de recepção dos equipamentos digitais instalados a partir do lançamento do SBTVD com os futuros formatos além da alta resolução 1080i. Estes equipamentos digitais já disponíveis serão denominados receptores legados e por continuidade de recepção entende-se que, se o conteúdo principal de um canal digital for transmitido em um novo formato, o mesmo conteúdo deve estar disponível simultaneamente para os receptores legados em formato e qualidade equivalentes aos praticados nas transmissões atuais.

Ao longo deste trabalho, devem ser cumpridos os seguintes objetivos específicos:

- Pesquisar e organizar a literatura relacionada à evolução da compressão e codificação de vídeo, realçando as ferramentas utilizadas no H.264/AVC e em sua extensão de escalabilidade. É necessário compreender os perfis inéditos deste padrão a fim de relacioná-los com a radiodifusão digital terrestre.

¹O formato de vídeo normalmente é indicado pelo número de linhas horizontais úteis e pelo tipo de seqüência de imagens (p: progressiva; i: entrelaçada). Dessa forma, 1080i representa 1080 linhas de vídeo entrelaçado.

- Identificar e apresentar as normas relevantes do SBTVD para a transmissão de vídeo, resumindo seus principais parâmetros, a fim de caracterizar o receptor legado e as possibilidades de evolução do sistema.
- Discutir a aplicação da escalabilidade de vídeo ao SBTVD e implantar um ambiente para testes de codificação e para testes sistêmicos.
- Realizar experimentos no ambiente para avaliar as características do sistema digital atual e das ferramentas utilizadas e para gerar referências para trabalhos futuros relativos à escalabilidade de vídeo no SBTVD.
- Consolidar as contribuições deste trabalho e o planejamento de trabalhos futuros.

1.2 Justificativa Histórica

A primeira transmissão de uma cena real para ser exibida remotamente através de um canal de comunicação foi realizada publicamente no dia 26 de Janeiro de 1926, em Londres, por John Baird, no aparato da Figura 1.1. Engenheiro escocês e responsável por contribuições essenciais ao primeiro sistema televisivo funcional, John Baird havia apresentado em seu laboratório no dia 2 de Outubro do ano anterior imagens em movimento do boneco ventríloquo Stooky Bill, com taxa de cinco figuras por segundo e 30 linhas verticais de resolução, compostas por tons de cinza. Na versão apresentada publicamente em 1926, o engenheiro já havia conseguido elevar a taxa para 12,5 figuras por segundo.



Figura 1.1: Aparato de transmissão do primeiro sistema televisivo de John Baird.

Fonte: <http://www.bairdtelevision.com/firstdemo.html>

Apesar de o telefone já dar ao homem a capacidade de ouvir e falar à distância na época de Baird, ver à distância representava um passo fundamental na emulação de capacidades humanas pelas telecomunicações (PEREIRA, 1997).

A fim de possibilitar essa experiência, diversos serviços de transmissão de vídeo com imagens em movimento surgiram ao longo dos anos. Contudo, o incrível avanço que houve na área deveu-se notadamente à rápida e ampla implantação e penetração mundial da televisão, que exerceu predominância como sistema audiovisual de comunicação durante o século passado.

Em pouco mais de 80 anos, após surgir como um sistema analógico desprovido de áudio e de cores, a televisão recebeu inúmeras melhorias funcionais. Inicialmente, ela evoluiu com a possibilidade de transmissão de áudio mono junto ao vídeo, depois de áudio estéreo, e ainda mais tarde de áudios adicionais em outros idiomas. Os sistemas televisivos evoluíram também com transmissões a cores, com aumento da taxa de quadros e, principalmente, com aumento de resolução. Quando a tecnologia permitiu adicionar dados digitais à transmissão ainda analógica, os sistemas puderam receber também alguns serviços adicionais como teletexto e legenda.

A predominância histórica da televisão teve como grande razão as características do sinal de vídeo (PEREIRA, 1997). Em comparação ao de áudio, por exemplo, o sinal de vídeo apresenta, usualmente, largura de banda relativamente grande, exigindo recursos muito significativos em termos de processamento, banda de transmissão e capacidade de armazenamento. Com isso, o tipo de transmissão da televisão, que ocorre de um ponto para muitos, proporciona melhor aproveitamento de recursos em comparação a outros serviços de caráter mais pessoal, realizados ponto a ponto.

A partir do final da década de 70, no entanto, o avanço da tecnologia de processamento, transmissão e armazenamento e das metodologias de representação dos sinais audiovisuais motivou a comunidade científica, que por sua vez contribuiu para uma rápida evolução no campo de compressão e codificação digital de vídeo. Surgiram assim as primeiras soluções multimídia no formato digital, ampliando-se os horizontes das comunicações. Com o tempo, serviços pessoais dessa natureza puderam ser progressivamente criados e disponibilizados para os usuários.

Atualmente, com a predominante digitalização dos canais de comunicação, existe uma grande quantidade de serviços e produtos multimídia acessíveis, incluindo celulares 3G, TV por IP, vídeo sob demanda, dispositivos pessoais móveis e portáteis com capacidade de armazenamento e reprodução de áudio e vídeo.

Neste contexto de diversidade, a necessidade de se produzir conteúdo para as distintas redes de comunicação e para os distintos serviços e dispositivos multimídia disponíveis e a necessidade de evolução destes sistemas, com algum grau de compatibilidade entre versões sucessivas, favorecem a consolidação de um padrão internacional para a codificação escalável de vídeo. Embora alguns outros padrões já tenham sido elaborados anteriormente, a falta de aplicações e serviços que impulsionassem esses padrões e a baixa eficiência das ferramentas de codificação escalável frente às ferramentas sem escalabilidade não colaboraram para que esses padrões se estabelecessem no mercado. Hoje, diferentemente, a indústria demonstra grande expectativa quanto à extensão de escalabilidade do H.264/AVC, vendo-a com bastante otimismo como um padrão de fato para o mercado (WIEGAND et al., 2007).

Em paralelo, voltando ao histórico da televisão, é importante observar também a evolução dos formatos de vídeo nos diversos sistemas que surgiram, pois esta evolução justifica a aplicação de escalabilidade proposta neste trabalho.

De forma geral, os formatos de vídeo a partir de certo momento foram especificados levando-se em consideração características do Sistema Visual Humano (*Human Visual System*) (HVS) além da disponibilidade tecnológica. Nos primórdios dos sistemas televisivos, a qualidade do vídeo era tão inferior à da visão normal de objetos naturais, que não era necessário considerar o HVS para justificar melhorias às especificações da televisão. Quando os sistemas passaram de 40 para 120 linhas, por exemplo, as principais considerações foram tecnológicas e econômicas. Por outro lado, o sistema denominado Televisão de Alta Definição (*High Definition Television*) (HDTV), que começou a ser estudado pelos japoneses já na década de 70, levou em consideração diversas características do HVS em sua especificação (SUGAWARA et al., 2008). Naquela ocasião, surgiram diversos trabalhos relacionando características do sistema de televisão às sensações do espectador, pois o intuito do HDTV era prover um novo formato com informações mais ricas, criando, de forma ampla, uma cultura visual mais rica. Foi somente com os sistemas digitais, no entanto, que este formato ganhou popularidade.

Pelo ganho expressivo de qualidade e de percepção que o HDTV proporciona, em um primeiro contato, os espectadores tendem a acreditar que este é o formato final em termos de evolução para apresentação de conteúdo visual. No entanto, sabe-se que tão logo os espectadores estejam acostumados a este formato, ainda haverá oportunidade de oferta de novos formatos superiores, causando novos impactos (SUGAWARA et al., 2008).

Pesquisas como a realizada pela empresa IBM (IBM INSTITUTE FOR BUSINESS VA-

LUE, 2006) indicam como principal referência para o futuro da televisão um usuário cada vez mais interativo. Interatividade, tradicionalmente, representa a possibilidade de interferência do usuário na programação/conteúdo, mas pode ser vista também como a sensação de imersão na programação. Dessa forma, o HDTV não deve encerrar a busca por sistemas de televisão com maiores resoluções.

A avaliação do HVS mostra que, mantendo-se a representação tradicional de vídeo por seqüências de imagens de duas dimensões, há possibilidade de evolução do HDTV. Porém, há um limite de saturação para a percepção de realidade que pode ser refletido em limites individuais na combinação de fatores como resolução do vídeo, taxa de quadros e ângulo de visão. Um limite excedido pode ocasionar desperdício de recursos ou até mesmo desconforto para o espectador. Como exemplos: pontos de uma matriz tornam-se indistinguíveis quando a distância entre eles é menor que um arco-minuto com relação à referência do observador; além disso, o preenchimento acima de uma dada porcentagem do campo de visão do espectador, em algumas situações, pode passar de aumento de percepção de realidade para sensação de enjôo.

Com essas informações, o centro de pesquisa da empresa japonesa NHK está trabalhando na especificação do sistema conhecido por Sistema de TV de Ultra Alta Definição (*Ultra High Definition TV*) (UHDTV), ou ainda Super Hi-Vision como foi mais recentemente denominado, que deverá operar nos limites da percepção de realidade do ser humano. Dentre os fatores que têm impacto sobre o HVS, alguns ainda devem ser estudados com maior profundidade, mas as influências da resolução e do campo de visão já se encontram bem avaliadas e documentadas por Sugawara et al. (2008). Por esse estudo, conforme mostra a Figura 1.2, a resolução horizontal do sistema deverá ser um múltiplo de 1920 até o limite de 7680, para um ângulo de visão horizontal de aproximadamente 100° , e a resolução vertical deverá ser um múltiplo de 1080 até o limite de 4320, para um ângulo de visão de aproximadamente 50° , configurando uma relação de aspecto da imagem de 16:9 e uma resolução até 16 vezes maior que a do HDTV.

Recentemente, a NHK já vem organizando demonstrações do sistema Super Hi-Vision no mundo operando no limite de resolução a uma taxa de 120 quadros por segundo. Nesta configuração, é possível dizer que o Super Hi-Vision não será compatível com os atuais sistemas de transmissão terrestre de televisão digital. Além de monitores e câmeras com a resolução especificada, uma das principais incompatibilidades será a banda necessária para transmissão. Dessa forma, espera-se que o processo completo de especificação e implantação do Super Hi-Vision tenha duração de cerca de 30 anos, assim como ocorreu com o sistema HDTV, e sua operação deve iniciar por fibra óptica e não pelo ar.

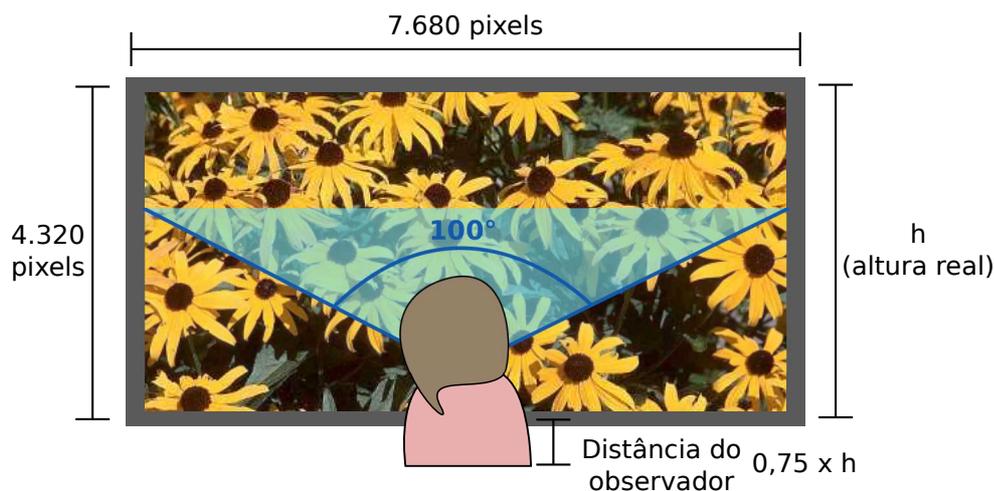


Figura 1.2: Representação parcial do sistema Super Hi-Vision

De qualquer forma, essa visão de longo prazo indica que, mesmo antes que um sistema como o Super Hi-Vision torne-se disponível para o usuário final, pode ser interessante evoluir os sistemas atualmente em operação no sentido de proporcionar maior imersão ao espectador, se houver a possibilidade tecnológica.

Atualmente os canais alocados para transmissão digital no Brasil suportam taxas da ordem de 20 Mbps e há expectativa de que a evolução dos algoritmos de codificação do H.264/AVC permita a transmissão de vídeo 1080i em ótima qualidade com taxas próximas à 10 Mbps. Dessa forma, suporte à codificação escalável pode ser uma solução tecnológica de curto e médio prazo que permita ao SBTVD transmissões de vídeo em formatos ainda não suportados, sem comprometer o legado principalmente de receptores do atual sistema.

1.3 Motivação

O Laboratório de Sistemas Integráveis (LSI) tem como objetivos principais desenvolver novas tecnologias, priorizando o interesse público e o desenvolvimento do país, e engajar seu pessoal técnico em programas de Mestrado e Doutorado (página de internet)². Especificamente o núcleo de Engenharia de Mídias deste laboratório, onde o autor desenvolve atividades de pesquisa e desenvolvimento há sete anos, vem trabalhando ao longo dos últimos 10 anos numa questão potencialmente alinhada a seus objetivos: a implantação da televisão digital no Brasil.

Um sistema de televisão digital é complexo e amplo, envolvendo diversas áreas de

²LSI. São Paulo. Apresenta o laboratório, seus núcleos, atividades e projetos. Disponível em: <http://www.lsi.usp.br/>. Acesso em: 18 ago. 2008.

conhecimento da engenharia elétrica. Ao longo dessa trajetória surgiram oportunidades de projetos acadêmicos, com o governo e com a iniciativa privada principalmente nas partes de recepção do sistema, com o desenvolvimento de terminais de acesso para o sistema digital, e de codificação multimídia.

O autor acredita que muitos resultados importantes já foram obtidos com os diversos projetos realizados e que outros tantos ainda virão, sendo esta sua principal motivação.

Em televisão digital, podem ser citados os seguintes trabalhos já realizados pelo núcleo de Engenharia de Mídias, nos quais o autor teve a oportunidade de contribuir direta ou indiretamente:

- Trabalhos de iniciação científica em unidade controladora de vídeo (relatório final)³ e em arquitetura de terminal de acesso (relatório final)⁴.
- Trabalho de conclusão de curso em 2004 dos alunos Laisa Caroline de Paula Costa e Rogério Pernas Nunes em decodificação de vídeo (relatório final)⁵.
- Doutorado de Marco Antônio Simon Dal Poz, defendido em 2005, intitulado: “Um terminal de acesso digital reconfigurável bidirecional adaptável aos padrões multimídia emergentes” (DALPOZ, 2005).
- Mestrado de Gil Barros, defendido em 2006, intitulado: “A consistência da interface com o usuário para a TV interativa” (BARROS, 2006).
- Mestrado de Rodrigo Admir Vaz, defendido em 2007, intitulado: “Uma interface de comunicação sem fio em TV digital baseada em rádio definido por programa de computador” (VAZ, 2007).
- Mestrado de Eduardo Carvalho, defendido em 2008, intitulado: “Uma plataforma modular para testes com interatividade na TV digital brasileira” (CARVALHO, 2008).
- Mestrado de Célio Hira, defendido em 2008, intitulado: “ARQUIMEDIA: Uma proposta de arquitetura de software para terminais de acesso à TV digital interativa” (HIRA, 2008).

³COSTA, L. C. de P. *Pesquisa e desenvolvimento de uma unidade controladora de vídeo reconfigurável baseada no padrão MPEG-4*. 2003. Projeto de Pesquisa - Escola Politécnica, Universidade de São Paulo, São Paulo. Apoiado pela FAPESP.

⁴NUNES, R. P. *Estudo, otimização e implementação de uma arquitetura reconfigurável para set-top-box digital*. 2003. Projeto de Pesquisa - Escola Politécnica, Universidade de São Paulo, São Paulo. Apoiado pela FAPESP.

⁵COSTA, L. C. de P.; NUNES, R. P. *Implementação de Blocos de Decodificação MPEG*. 2004. Projeto de Conclusão de Curso - Escola Politécnica, Universidade de São Paulo, São Paulo.

- Projeto do Terminal de Acesso de Referência (TAR) do SBTVD em 2005.
- Projetos de codificação de áudio AAC e vídeo H.264/AVC para o SBTVD em 2005.
- Participação nos grupos técnicos de especificação do SBTVD.
- Projetos de terminais de acesso em conformidade com o SBTVD para a iniciativa privada.

1.4 Relevância

O Decreto nº 4.901 (BRASIL, 2003), que instituiu o SBTVD em 26 de Novembro de 2003 como um instrumento sócio-cultural de inclusão, também deixa claro os objetivos de desenvolver a tecnologia no país, aprimorando qualidade de áudio e vídeo e aperfeiçoando o uso do espectro de radiofrequências, estimulando também a evolução das exploradoras dos serviços de televisão.

Dessa forma, os resultados esperados a partir dos objetivos estabelecidos neste trabalho estão alinhados com o contexto da sociedade brasileira. Além disso, estes resultados, incluindo a síntese bibliográfica sobre a área de codificação de vídeo, devem proporcionar ao LSI a oportunidade de orientar novos trabalhos de mestrado, doutorado e iniciação científica na área e a consolidar novos projetos com o setor produtivo, colaborando com o engajamento de novos pesquisadores nos programas de pós-graduação e com o desenvolvimento do país.

1.5 Estrutura da Dissertação

Este texto está dividido em seis capítulos. O primeiro é esta introdução e os demais estão alinhados aos objetivos específicos propostos:

O Capítulo 2, especialmente dedicado à escalabilidade de vídeo, introduz rapidamente os conceitos de compressão e codificação de vídeo (material de suporte para os fundamentos da representação de vídeo analógico e digital podem ser encontrados no Apêndice A); em seguida, apresenta a evolução histórica das ferramentas e normas de compressão e codificação de vídeo, conduzindo aos recentes desafios da área e às ferramentas para escalabilidade. Na seqüência, apresenta em destaque a norma H.264/AVC, que é a base para a extensão de suporte à escalabilidade mais promissora para a indústria (SVC). Por fim, a extensão SVC é apresentada em detalhes para fechar o capítulo.

O Capítulo 3 apresenta o histórico de trabalhos que conduziram à especificação e implantação do Sistema Brasileiro de TV Digital Terrestre (SBTV-D-T) e os resultados destes trabalhos, ressaltando as partes relativas à transmissão, à camada de transporte de informação do sistema e à codificação de vídeo.

O Capítulo 4 apresenta considerações sobre a escalabilidade de vídeo para o SBTV-D, considerando a norma H.264/AVC e sua extensão de escalabilidade. O capítulo descreve projetos correlatos e discute a evolução dos formatos de vídeo no sistema. Por fim, apresenta também um ambiente experimental para avaliações sobre a codificação escalável.

O Capítulo 5 concentra resultados experimentais preliminares, obtidos no ambiente especificado e implementado neste trabalho. Dentre os resultados são evidenciadas premissas importantes da codificação escalável SVC.

O Capítulo 6 apresenta as conclusões finais, com considerações sobre o atendimento dos objetivos propostos e com a discussão dos trabalhos futuros, derivados dos resultados obtidos e análises realizadas neste trabalho.

2 Codificação Escalável de Vídeo

Principalmente ao longo dos últimos 20 anos, as tecnologias de compressão digital de vídeo tornaram-se parte integrante do modo como criamos, transmitimos e consumimos informação visual. São inúmeros os exemplos de cenários que hoje se beneficiam da compressão digital de vídeo, tais como: radiodifusão aberta, serviços de televisão por assinatura através de cabo ou satélite, serviços de comunicação bi-direcional em tempo real, transmissão de fluxo de vídeo por internet ou redes locais, armazenamento de conteúdo, transmissão de mensagens multimídia por rede de telefonia (SULLIVAN; WIEGAND, 2005).

O ritmo acelerado de desenvolvimento em compressão e codificação de vídeo deve-se, em grande parte, à sistematização criada por grupos importantes de especialistas que surgiram nos órgãos internacionais de padronização como o grupo *Moving Picture Experts Group* (MPEG) dentro da ISO/IEC e o grupo *Video Coding Experts Group* (VCEG) dentro da ITU-T. Estes grupos concentraram esforços na criação de normas internacionais que visam garantir a interoperabilidade de sistemas de vídeo digital mantendo a flexibilidade para adaptar as ferramentas de compressão e codificação aos requisitos específicos de cada aplicação (LUTHRA; SULLIVAN; WIEGAND, 2003). Normas de compressão e codificação de vídeo como ITU-T H.261, ISO/IEC MPEG-1, ITU-T H.262 ou ISO/IEC MPEG-2 Video, ITU-T H.263, ISO/IEC MPEG-4 Visual e ITU-T H.264 ou ISO/IEC AVC tornaram-se amplamente utilizadas na indústria.

A evolução das ferramentas consolidada nas normas citadas sempre buscou balancear parâmetros como taxa máxima de bits do canal ou dispositivo de armazenamento, distorção do vídeo decodificado, atraso máximo entre captura e recuperação do conteúdo e complexidade computacional do sistema de compressão e codificação. O problema básico que cada norma busca resolver para uma ou mais aplicações é possibilitar máxima fidelidade de qualidade do conteúdo recuperado depois da codificação em comparação ao material original ou mínima taxa de bits possível para uma determinada fidelidade de reprodução (SULLIVAN; WIEGAND, 2005).

Naturalmente, com a contínua evolução das ferramentas de compressão e codificação,

a tendência é que a variada gama de cenários de aplicações de vídeo digital continue crescendo. Serviços baseados em conteúdo tridimensional podem ser citados como exemplos de aplicações das novas tecnologias de vídeo, e num futuro bem próximo, devem integrar equipamentos do dia-a-dia.

Além da ampliação dos cenários de uso de vídeo digital, outra tendência marcante no âmbito da compressão e codificação é de que a produção de conteúdo possa ser feita de tal forma que sua distribuição seja transparente, ou seja, deseja-se que uma vez produzido, o mesmo conteúdo possa ser consumido na maior quantidade possível de cenários distintos. Dessa forma, do ponto de vista do fornecedor, o custo de produção e distribuição diminui e do ponto de vista do consumidor, o acesso ao conteúdo desejado é facilitado, o que colabora para a convergência e alta disponibilidade de conteúdo.

Ferramentas de escalabilidade recentemente desenvolvidas, além de virem de encontro a essa necessidade, permitem também que sistemas em operação possam ter melhorias de qualidade sem comprometer os equipamentos legados. Ou seja, é possível melhorar a qualidade do vídeo para novos equipamentos de um sistema sem que os equipamentos antigos sejam prejudicados, de maneira análoga à introdução da televisão a cores, que não impediu a recepção dos televisores ainda em preto e branco usando o mesmo canal.

Para melhor compreender as características de estado-da-arte da norma H.264/AVC e de sua extensão de escalabilidade, que é o tema central deste capítulo, é necessário antes uma boa compreensão da representação de vídeo digital apresentada no Apêndice A e de conceitos básicos de compressão e codificação de vídeo, apresentados adiante.

2.1 Compressão e Codificação Digital

A compressão digital de vídeo é o mecanismo utilizado na redução da informação total necessária para representar o conteúdo original através de técnicas descritas por algoritmos que se baseiam em dois princípios fundamentais: redução de redundâncias e redução de irrelevâncias do conteúdo. Os algoritmos de redução de redundâncias caracterizam uma compressão sem perdas e são totalmente reversíveis. Por outro lado, os algoritmos de redução de irrelevâncias, que são subjetivas e dependem de características psicofisiológicas do ser humano, caracterizam uma compressão com perdas, a partir da qual a informação original não pode mais ser recuperada na íntegra. A medida da degradação por perdas é referenciada como taxa de distorção e quando exagerada significa resultados perceptíveis ou até inaceitáveis na qualidade do conteúdo comprimido. Para a maior

parte das aplicações de vídeo, no entanto, a degradação é necessária em algum grau para melhor aproveitamento de recursos.

A codificação digital, por sua vez, está associada à representação de blocos de informação por símbolos pré-definidos. Sistemas simples de comunicação como código Morse e posteriormente Modulação por Código de Pulsos (*Pulse Code Modulation*) (PCM) já são exemplos de sistemas com codificação digital.

Estes dois mecanismos, compressão e codificação, que já nos anos 80 e principalmente 90 passavam a integrar os sistemas de vídeo, não possuem aplicação de forma isolada, pois dependem de seus mecanismos inversos: a decodificação, para reverter o processo de codificação e a descompressão, necessária para reverter o processo de compressão.

Na prática a compressão e a codificação aparecem sempre integradas, o que faz com que na literatura seja muito comum a referência a apenas uma delas quando se deseja, na verdade, referenciar ambas. Neste texto, por simplicidade, também será utilizada apenas *compressão* ou *codificação* para todo o processo de compressão e representação digital, dependendo do contexto e da ênfase desejada. Também será utilizada apenas *descompressão* ou *decodificação* com o mesmo propósito.

Ao conjunto de compressores/codificadores e decodificadores/descompressores de um sistema multimídia, tanto para vídeo quanto para áudio, dá-se o nome de CODEC.

Em cadeias típicas de sistemas multimídia, observam-se alguns fluxos alternativos da informação desde a captura até o consumo final, podendo haver emprego de CODECs apenas no fluxo de transmissão, no fluxo de armazenamento para edição, ou ainda no armazenamento para consumo. É muito comum a informação ser recodificada entre estágios distintos, sendo essa operação de transformação conhecida como transcodificação.

CODECs que processam apenas algoritmos sem perdas proporcionam uma taxa de redução de informação tipicamente entre 2:1 e 5:1, enquanto que os que processam algoritmos com perdas conseguem proporcionar atualmente, ainda com qualidade satisfatória, redução típica entre 50:1 e 100:1.

2.2 Análise de Qualidade de CODECs de Vídeo

Conforme Sullivan e Wiegand (2005), CODECs de vídeo são primeiramente classificados em termos da taxa de bits do fluxo codificado para o canal e da distorção do vídeo decodificado, onde distorção tem o significado mais geral de qualidade. Além disso, consideram-se

também a latência de início de apresentação na decodificação (atraso ponta a ponta do sistema) e a complexidade computacional dos algoritmos utilizados.

Com base nessas características, o problema prático de projeto de um CODEC pode ser resumido como: a partir de valores máximos determinados para o atraso do sistema e para a complexidade computacional, busca-se um compromisso ótimo entre a taxa de bits e a qualidade do CODEC para o conjunto de aplicações vislumbradas.

Dessa forma, medir qualidade de vídeo é fundamental para que se possa especificar, avaliar e comparar os diversos sistemas de comunicação.

Medir qualidade de vídeo, no entanto, não é uma tarefa fácil, pois além de muitos fatores poderem influenciar nos resultados, boa parte deles são intrinsecamente de natureza subjetiva. A percepção de qualidade pode variar, por exemplo, de acordo com a tarefa que o espectador está executando, como assistir passivamente a um vídeo, participar de uma videoconferência, ou ainda identificar uma pessoa num sistema de monitoramento de segurança. A percepção pode variar ainda de acordo com fatores culturais, principalmente em relação a tons.

Medidas objetivas de qualidade de vídeo existem e oferecem resultados precisos e reprodutíveis. Algumas delas simulam os fatores subjetivos através de modelos sofisticados do HVS e da percepção, mas até o momento não há um conjunto de medidas objetivas que substitua por completo a análise subjetiva no julgamento de qualidade.

Em Outubro de 1997, durante um encontro de especialistas em qualidade de vídeo, foi criado um grupo com participantes de diversas nacionalidades, a maior parte pertencente a grupos de estudos da ITU, com a finalidade de endereçar as questões relacionadas à área. Este grupo, denominado *Video Quality Experts Group* (VQEG)¹, vem trabalhando na definição de metodologias para avaliar as ferramentas de medida objetiva de qualidade de vídeo em comparação às medidas subjetivas, realizadas com base na recomendação ITU-R BT.500, que é a principal referência para análise desse tipo.

Pelo cronograma do grupo VQEG (VQEG HDTV GROUP, 2008), os testes com ferramentas objetivas de análise de conteúdo em alta definição serão realizados ao longo de 2009, após a criação de um acervo de vídeo de referência e da coleta das propostas de ferramentas.

Dentre as medidas objetivas conhecidas, uma das mais populares é a Razão Pico Sinal-Ruído (*Peak Signal-to-Noise Ratio*) (PSNR), apresentada na eq. (2.1), onde N é o

¹Página da internet oficial do grupo, disponível em: <http://www.vqeg.org>. Acesso em: 02 dez. 2008.

número de bits de cada amostra. Apesar de não ser a mais sofisticada, esta medida é popular por sua rapidez e simplicidade de cálculo. Ela é realizada em escala logarítmica em função do Erro Quadrático Médio (*Mean Squared Error*) (MSE). O MSE por sua vez é obtido pela eq. (2.2), entre duas imagens I e K monocromáticas de dimensão $m \cdot n$, sendo uma a imagem original e outra a imagem sob avaliação.

$$\text{PSNR} = 10 \cdot \log_{10} \frac{(2^N - 1)^2}{\text{MSE}} \quad (2.1)$$

$$\text{MSE} = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I(i, j) - K(i, j))^2 \quad (2.2)$$

Muitos resultados de ferramentas de compressão são apresentados na literatura com valores de PSNR, que serve bem para uma primeira análise de eficiência e para avaliações comparativas entre configurações de um mesmo CODEC.

Recentemente foi desenvolvido um novo método de avaliação de qualidade de imagens, denominado Similaridade Estrutural (*Structural SIMilarity*) (SSIM). Baseado na hipótese de que o HVS é altamente adaptado à extração de informação estruturada do campo visual, concluiu-se que deveria haver alguma medida de similaridade estrutural entre imagens capaz de prover boa aproximação de sua qualidade percebida (WU; RAO, 2005). Enquanto métodos objetivos clássicos de avaliação de qualidade procuravam explorar propriedades perceptivas elementares do HVS, o SSIM partiu de uma constatação macroscópica das características do HVS.

A medida de similaridade estrutural entre imagens é refletida em um índice, que pode ser calculado de diversas formas, explorando diferentes estruturas das imagens, ou atribuindo pesos distintos para a distorção medida em cada estrutura. Um exemplo de cálculo de índice SSIM foi descrito por Wang et al. (2004), conforme eqs. (2.3) e (2.4). Nestas, \mathbf{x} e \mathbf{y} são sub-regiões espacialmente equivalentes das duas imagens, μ_x e μ_y são as intensidades médias das sub-regiões, σ_x e σ_y são os desvios padrões das intensidades nas sub-regiões, σ_{xy} é a covariância entre \mathbf{x} e \mathbf{y} , C_1 e C_2 são constantes e MSSIM é o índice médio das imagens \mathbf{X} e \mathbf{Y} completas, considerando todas as M sub-regiões em que estas foram divididas para análise.

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

$$\text{MSSIM}(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{j=1}^M \text{SSIM}(\mathbf{x}_j, \mathbf{y}_j) \quad (2.4)$$

Tabela 2.1: Efeitos comuns em compressão e codificação de vídeo

Efeito	Definição
Efeito de Bloco	Distorção da imagem caracterizada pelo aparecimento de padrões retangulares ou tipo tabuleiro não apresentados na imagem original
<i>Blurring</i>	Uma distorção global sobre a imagem, caracterizada pela perda de resolução de bordas e detalhes
<i>Edge Busyness</i>	Distorção concentrada perto das bordas dos objetos
<i>Jerkiness</i>	O plano original e o movimento contínuo é percebido como uma série de quadros bruscos (degradação temporal)
Erro de bloco	Uma forma de distorção de bloco onde um ou mais blocos da imagem não possui semelhança com a cena corrente. É causado por erros introduzidos pelo canal de transmissão
Ruído Tipo Mosquito	Ruído de quantização caracterizado por artefatos se movendo ao redor de bordas (como um mosquito voando ao redor da cabeça e ombros de uma pessoa)
Degradação de resposta ao movimento	Deterioração de movimento de vídeo de forma que imagem tenha uma perda de resolução espaço-temporal
Persistência de objeto	Um objeto que foi apagado e continua a aparecer no sinal de imagem recebido
Reposta de corte de cena	Degradação associada ao corte da cena
<i>Smearing</i>	Uma distorção localizada acima da subregião da imagem recebida, caracterizada por redução da definição de bordas e detalhes espaciais

A Tabela 2.1 apresenta alguns dos efeitos mais comuns observados em vídeos digitais que passaram pelo processamento de um CODEC. Dentre os efeitos listados, o *efeito de bloco* é normalmente o mais perceptível, e, por isso, é também o efeito que os fabricantes de soluções de compressão procuram amenizar com mais afinco.

2.3 Evolução dos CODECs de vídeo

As principais técnicas ou ferramentas de compressão digital de vídeo que surgiram no século passado foram, inicialmente, estudadas de forma independente. Pelas restrições de tecnologia, elas eram soluções auto-suficientes e de certa forma vistas até mesmo como soluções concorrentes entre si. Somente o tempo e a evolução tecnológica permitiram que essas diversas técnicas pudessem ser usadas simultaneamente em uma mesma solução (READER, 2002), aumentando a eficiência de codificação.

A Figura 2.1 apresenta as principais ferramentas que surgiram entre as décadas de 50 e 80 e que serviram de base para a evolução dos CODECs de vídeo.

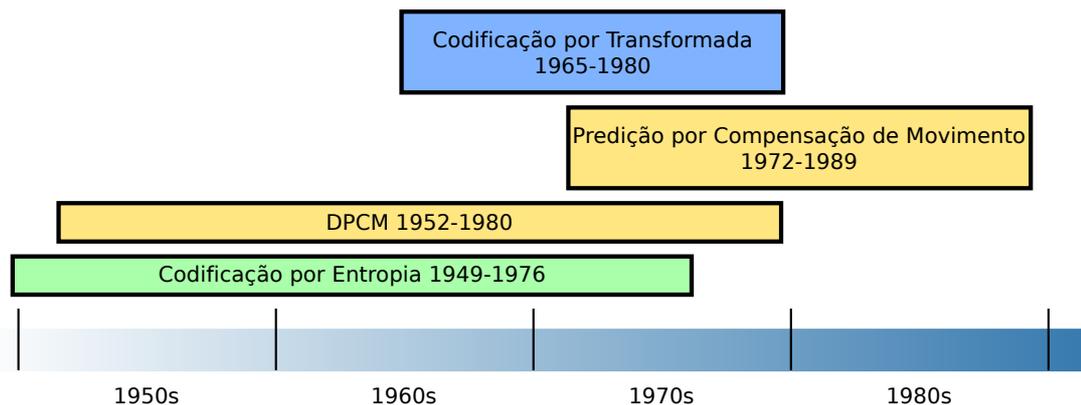


Figura 2.1: Principais técnicas de compressão digital dos anos 50 aos 80

Cada ferramenta está acompanhada de duas datas. A de início refere-se à data da primeira publicação relacionada ao tema e a de término refere-se a uma aproximação de quando os desenvolvimentos já haviam atingido grau de maturação suficiente para que a ferramenta fosse incorporada a uma norma internacional.

As características das ferramentas apresentadas podem ser resumidas como:

Codificação por entropia: criada para remover as redundâncias estatísticas dos códigos que representam uma imagem ou vídeo, a codificação por entropia associa símbolos menores a seqüências mais freqüentes. Pode ser aplicada com ou sem o uso prévio de outras ferramentas.

DPCM: a Modulação Diferencial por Código de Pulsos (*Differential Pulse Code Modulation*) (DPCM) surgiu somente para redução de redundância espacial dentro de uma mesma imagem.

Predição por compensação de movimento: pode ser considerada uma evolução da DPCM, por isso está representada com a mesma cor na figura. Como era sabido desde 1929, existe elevada correlação temporal entre imagens em conteúdo de vídeo. Além da redundância espacial, a temporal também passou a ser explorada.

Codificação por transformada: realizada em blocos de pixels com tamanhos adaptáveis, a transformada matemática procura descorrelacionar os valores dos pixels, gerando um novo conjunto de amostras no domínio das freqüências, onde a redundância espacial é removida pela compactação da energia dos coeficientes espaciais em alguns poucos coeficientes da transformada.

Baseado apenas em modulação por DPCM e substituição condicional de áreas das figuras codificadas, o primeiro CODEC de vídeo padronizado na história foi o COST211

em 1984 pela CCITT. Resultado das técnicas de codificação digital que se consolidavam, da baixa eficiência da digitalização PCM disponível, da motivação por viabilizar sistemas de vídeo e da disponibilidade de tecnologia mais avançada, o COST211 foi padronizado com a referência de H.120 na CCITT. Apesar de boa resolução espacial, apresentava qualidade temporal muito baixa.

De qualquer forma, a iniciativa de transformar um CODEC de vídeo em uma norma internacional aberta foi seguida posteriormente pela ITU e pela ISO/IEC que foram os principais órgãos de normalização de CODECs após o H.120. Esse processo de normalização mostrou suas vantagens, pois a garantia de interoperabilidade entre equipamentos de fabricantes diferentes, além de facilitar o uso, permite ampla adoção de um padrão pela indústria e conseqüente redução de custos pela escala da produção.

Outras duas características importantes das normas publicadas pela ITU e pela ISO/IEC são a liberdade de implementação dos codificadores e a assimetria de complexidade computacional entre codificadores e decodificadores. Uma vez que as normas definem apenas a sintaxe permitida para um fluxo de bits codificado e o processo de decodificação, não há imposição sobre as técnicas utilizadas na implementação dos codificadores, desde que gerem fluxos de bits em conformidade com as normas. Com isso, fabricantes podem criar diferenciais competitivos em função de algoritmos mais sofisticados. Além disso, as normas priorizam baixa complexidade computacional dos decodificadores em relação aos codificadores. Dessa forma, como a maior parte dos sistemas é formada por um codificador alimentando múltiplos decodificadores, o produtor de conteúdo arca com maiores custos para desonerar os diversos consumidores.

Normalmente, codificação por transformada na literatura refere-se à ampla família de técnicas de codificação que envolvem a associação da transformada propriamente dita com a codificação por entropia. A classe dos codificadores híbridos baseados em Transformada Discreta do Cosseno (*Discrete Cosine Transform*) (DCT) e a classe dos codificadores baseados em transformadas por *wavelets* pertencem a esta família e serão tratadas adiante.

A Figura 2.2 apresenta a seqüência de CODECs padronizados pela ITU e pela ISO/IEC ao longo do tempo. É possível observar, além das normas que foram desenvolvidas exclusivamente por uma das duas organizações, as normas que foram definidas em parceria, através da união de esforços dos seus grupos de especialistas. É possível observar também as normas relacionadas à codificação de imagens estáticas desenvolvidas pelo grupo JPEG da ISO/IEC. Estas normas compartilham algumas ferramentas espaciais (de codificação intra-quadro) com as normas de vídeo.

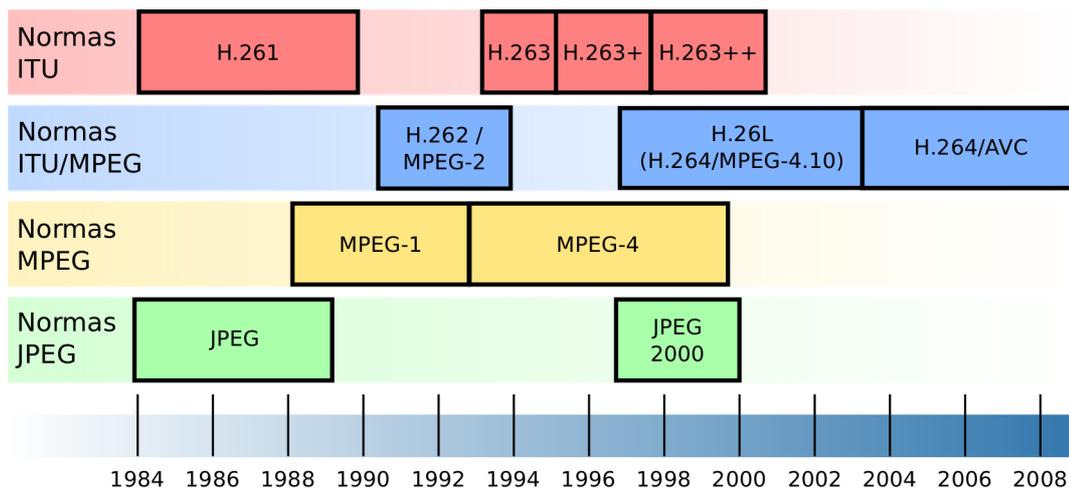


Figura 2.2: Normas de codificação de vídeo publicadas pela ITU e pela ISO/IEC

2.3.1 Codificadores Híbridos ISO/ITU

A norma do H.261 foi a primeira a consolidar as três principais ferramentas de compressão digital apresentadas anteriormente, especificando o primeiro CODEC híbrido da história. A estrutura deste CODEC foi essencialmente mantida nas normas posteriores, que basicamente agregaram novas ferramentas e aprimoraram as já existentes.

Os codificadores híbridos recebem este nome devido à associação da predição por compensação de movimento com a codificação por transformada, misturando, assim, o domínio do espaço e do tempo com o domínio da frequência espacial resultante da transformada.

De forma resumida, o histórico das principais normas de codificadores de vídeo híbridos da ITU e da ISO/IEC até o momento foi o seguinte: com início em 1984, a norma do H.261 teve seus trabalhos substancialmente concluídos em 1989, ainda pela organização CCITT, que posteriormente passou para ITU. Esta norma, como mencionado, foi a precedente direta das normas MPEG. MPEG-1 e H.261 compartilham estruturas de dados, ferramentas de codificação e elementos sintáticos comuns, mas ainda assim, o MPEG-1 não é retro compatível com o H.261, embora possa ser visto como um superconjunto dele. O desenvolvimento do MPEG-1 começou em 1988, sendo significativamente concluído em 1992. O H.262/MPEG-2, que pode ser visto como um superconjunto do MPEG-1, começou em 1990, sendo substancialmente concluído em 1994. O H.263 começou em 1992 e teve sua primeira versão publicada em 1995. O MPEG-4, construído a partir do H.262/MPEG-2 e do H.263, foi iniciado em 1993 e sua primeira versão foi concluída em 1998. Por fim, o desenvolvimento do H.26L começou em 1998, e a primeira publicação

conjunta entre ITU e ISO/IEC, da norma resultante, ocorreu em 2003. A norma foi identificada como H.264 na ITU e MPEG-4 Parte 10 na ISO/IEC e atualmente já houve a publicação de novas versões desta norma como será descrito mais adiante.

Reader (2002) apresenta de forma bastante completa as dezenas de artigos científicos mais significativos na história da evolução dos codificadores híbridos e as relações entre eles. Além disso, relaciona também os motivos e as escolhas que levaram à definição da estrutura híbrida dos codificadores, apresentada genericamente na Figura 2.3.

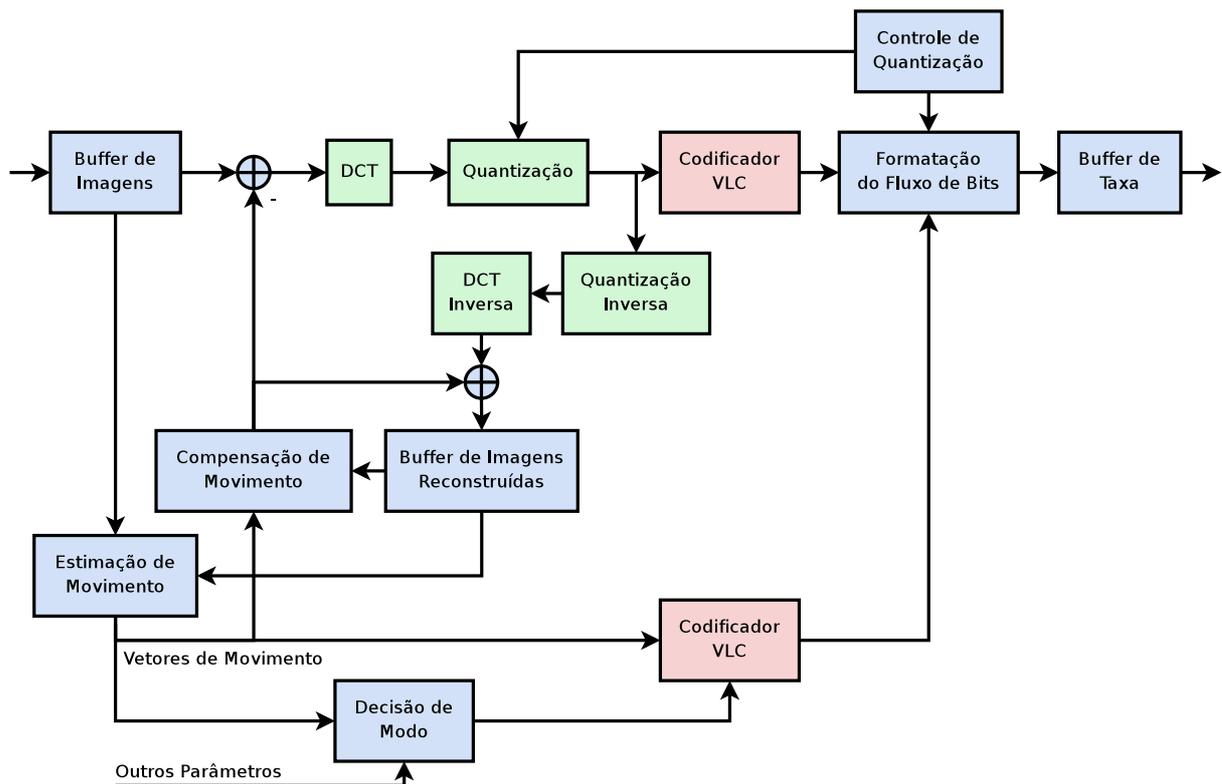


Figura 2.3: Estrutura genérica de um codificador híbrido

Nesta estrutura, a seqüência de imagens já digitalizadas do vídeo (a imagem é composta por matrizes de luma e croma neste estágio) é parcialmente acumulada no buffer de entrada do codificador. Cada imagem deste buffer é lida em blocos bi-dimensionais de pixels, geralmente no formato 16x16 (conhecidos como macroblocos). Os blocos das imagens que serão codificadas apenas com as ferramentas espaciais (imagens intra-quadro, referenciadas como imagens tipo I, e que, geralmente, iniciam um grupo de imagens da seqüência) passam pelo primeiro somador sem serem alterados, seguindo para a ferramenta DCT, da qual resultam blocos de mesma dimensão com amostras das freqüências espaciais dos blocos originais. Estas amostras são então quantizadas e em seguida codificadas pela ferramenta Código de Comprimento Variado (*Variable Length Code*) (VLC),

que diminui a entropia dos códigos. Por fim, a codificação da imagem é colocada em forma de fluxo de bits para transmissão com assessoria de um controlador de taxa que atua também nos parâmetros de quantização, alterando a qualidade das imagens codificadas de acordo com a disponibilidade do buffer de saída. Cada imagem codificada passa pelo processo de decodificação (fluxo inferior da figura) para alimentar as ferramentas que são utilizadas na codificação de imagens com predição e estimação de movimento (imagens inter-quadro, referenciadas como tipo P quando dependem apenas de uma imagem pré-codificada e tipo B quando dependem de duas). Dos blocos das imagens que são codificadas com predição de movimento, entram apenas os resíduos da diferença com os blocos de referência na DCT e em paralelo os vetores que deverão ser utilizados para a compensação de movimento na decodificação são também tratados por um bloco VLC para compor o fluxo de bits final.

A Figura 2.4 apresenta a evolução de eficiência, e, por conseqüência, de qualidade, de codificação das normas descritas anteriormente, evidenciando as principais características das ferramentas responsáveis pelo ganho em comparação às normas precedentes.

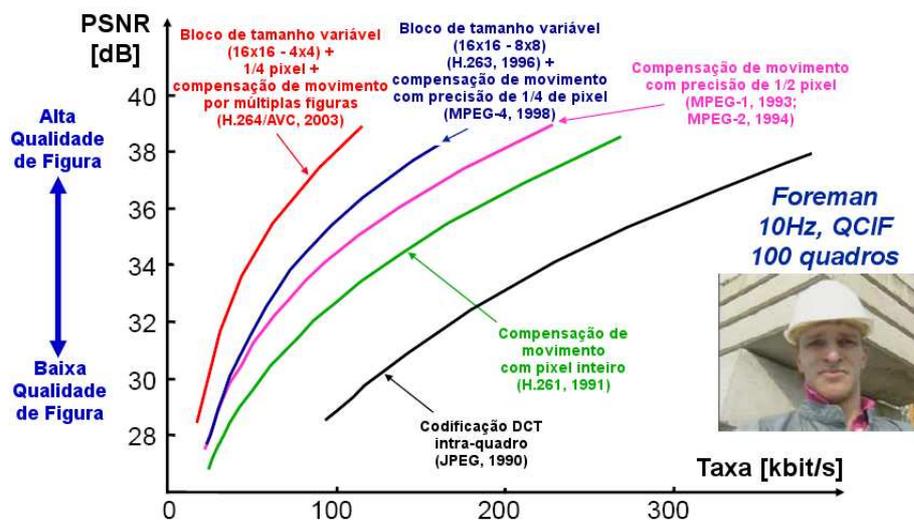


Figura 2.4: Marcos da codificação de vídeo. Fonte: ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/netmedia/20080416-17-garysull-isce_en.pdf

É possível observar o ganho expressivo de eficiência de codificação do H.264/AVC em relação aos demais codificadores. Com a seqüência de vídeo testada, cujas características estão apresentadas ao lado direito, para uma qualidade de 38 dB, por exemplo, obtém-se com o H.264/AVC uma taxa de 100 kbps, enquanto com o MPEG-2, obtém-se o dobro de taxa, 200 kbps. Analogamente, para uma taxa de 100 kbps, com o H.264/AVC obtém-se a qualidade de 38 dB, enquanto com o MPEG-2 obtém-se apenas 35 dB.

2.3.2 Codificadores Baseados em *Wavelets*

A norma H.264/AVC e suas extensões não apresentam suporte à classe dos codificadores baseados em *wavelets*. Nesta seção será apresentada apenas uma breve contextualização do tópico e algumas referências que podem ser exploradas para sua melhor compreensão, pois na parte de escalabilidade, este tipo de codificação será novamente citada.

Wavelets são funções definidas sobre um intervalo finito e que apresentam correlação cruzada nula. A idéia básica da transformada *wavelet* é representar qualquer função $f(t)$ arbitrária como a sobreposição de um conjunto de funções do tipo *wavelet*. Esta sobreposição decompõe $f(t)$ em diferentes níveis de escala, sendo cada nível decomposto novamente com uma resolução adaptada. As funções de base para decomposição são obtidas a partir de um único protótipo básico de função *wavelet* por redimensionamentos e deslocamentos.

A transformada *wavelet* pode ser considerada um caso especial de transformada por sub-bandas. A análise de decomposição por sub-bandas está geralmente focada no domínio da frequência, enquanto na transformada *wavelet* tanto as características espaciais quanto de frequências do sinal devem ser consideradas (DAVIS; NOSRATINIA, 1998).

A representação por *wavelets* é particularmente bem adaptada para a captura tanto de transientes de alta frequência (bordas e cantos) quanto de extensas áreas espaciais de baixa frequência (fundos) (DANYALI, 2004).

O modo prático e rápido de se implementar transformadas do tipo *wavelet* é através de bancos de filtros, sendo que esta relação foi primeiro descoberta por Daubechies (1988). A complexidade computacional e a simetria entre codificador e decodificador deste tipo de solução, no entanto, não a tornaram tão atrativa comercialmente até o momento.

Das normas apresentadas anteriormente, na Figura 2.2, o JPEG-2000 é um exemplo de codificação baseada em transformada *wavelet*, mas trata-se de um CODEC para imagens estáticas. O MPEG-4, por sua vez, embora seja substancialmente baseado em DCT, apresenta suporte à transformada *wavelet* para texturas em um dos seus perfis.

2.3.3 Alguns Trabalhos Atuais

Atualmente está sendo desenvolvido um CODEC aberto baseado em transformada *wavelet*, cujo nome é Dirac². Este CODEC junta técnicas de compensação de movimento por

²Detalhes deste CODEC fogem do escopo deste trabalho, mas podem ser obtidos na página de internet oficial do projeto, disponível em <http://diracvideo.org/>. Acesso em: 10 dez. 2008.

blocos sobrepostos com a transformada *wavelet* e tem recebido cada vez mais atenção da comunidade científica e da indústria.

Uma outra técnica importante que continua sendo desenvolvida é a formação de cenas por composição de objetos com formatos arbitrários. Esta técnica não é recente, pois já fazia parte da norma MPEG-4 Visual, mas continua atual no sentido de expectativa de se tornar amplamente difundida em aplicações de vídeo. Na representação de cenas por objetos, cada objeto é codificado individualmente com representação própria, podendo ser usada a representação planar tradicional ou outra mais adequada ao tipo de objeto (como representações especiais de imagens sintéticas, por exemplo).

De forma geral, no contexto da codificação de vídeo, vale citar ainda algumas outras linhas pesquisadas atualmente, que incluem: outros métodos de compressão como compressão por sub-bandas e composição de transformadas com técnicas de compensação de movimento; algoritmos que facilitam a transcodificação de conteúdo entre CODECs distintos; codificação multicâmera utilizada para vídeo 3D e para espectador com ponto de observação livre; codificação reconfigurável, que permitirá a composição de codificadores e decodificadores pela conexão de ferramentas de um arcabouço em tempo de execução; CODECs desbalanceados com maior complexidade computacional no decodificador; CODECs distribuídos; e finalmente, técnicas e ferramentas para a escalabilidade de vídeo, que é o tema central deste capítulo.

2.4 Escalabilidade de Vídeo

Conforme já descrito na Introdução com outras palavras, escalabilidade de vídeo no contexto deste trabalho pode ser entendida como um formato de codificação que permita a geração do fluxo de bits de um vídeo codificado uma única vez, enquanto capaz de ser manipulado para extração de múltiplas representações do conteúdo original através da decodificação de apenas partes do fluxo de bits ou de todo ele.

Para se obter esse formato de codificação, durante muito tempo buscou-se o desenvolvimento de ferramentas de escalabilidade que pudessem ser combinadas com os modelos tradicionais de codificadores híbridos. Foi o conceito de *codificação por camadas*, originalmente proposto por Ghanbari (1989) para melhorar a robustez de CODECs de vídeo contra perdas de pacotes em redes ATM, que serviu de base para esses trabalhos.

A Figura 2.5 mostra genericamente o esquema de codificação por camadas, no qual uma *camada base* é codificada contendo apenas informação que representa uma versão

reduzida do conteúdo original, seja em termos de taxa de quadros, resolução ou qualidade. Uma ou mais *camadas de enriquecimento* são então codificadas e adicionadas para melhorar progressivamente as características do vídeo. A rede de comunicação e/ou os receptores, por fim, podem selecionar a camada que será transmitida e reproduzida, sendo que uma determinada camada só pode, necessariamente, ser decodificada juntamente com as informações das camadas inferiores.

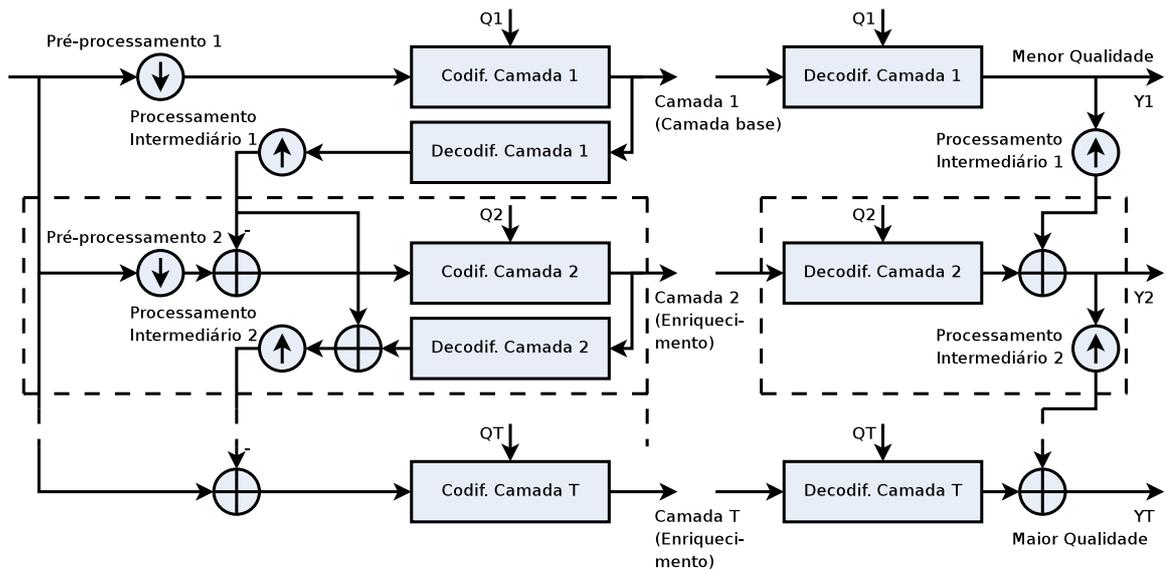


Figura 2.5: Esquema de codificação escalável por camadas

Enquanto as primeiras normas de codificação de vídeo como H.261 e MPEG-1 não proviam qualquer mecanismo de escalabilidade, pela orientação a aplicações específicas, a H.262/MPEG-2 foi a primeira norma de codificação de vídeo de uso geral a incluir ferramentas de suporte à escalabilidade, em função do desejo de retro compatibilidade com o MPEG-1. Esperava-se que informação básica do fluxo de bits pudesse, eventualmente, ser codificada e decodificada com o padrão antigo enquanto informação extra para melhoria de qualidade fosse processada pela nova norma.

Dessa forma, muitas das ferramentas desenvolvidas para o esquema de codificação por camadas foram incorporadas inicialmente à norma MPEG-2, que apesar de ter sido limitada ao máximo de três camadas para qualquer perfil de uso, já contempla os três tipos principais de escalabilidade que serão descritos adiante: *espacial*, *temporal* e *Razão Sinal-Ruído* (Signal-to-Noise Ratio) (*SNR*) ou *de qualidade*. Além disso, a norma do MPEG-2 ainda suporta particionamento de dados (*data partitioning*), que permite a separação do fluxo de bits em camadas distintas, de acordo com a importância dos elementos de codificação carregados para a qualidade do sinal reconstruído.

Após a norma do MPEG-2, ferramentas ainda mais flexíveis de escalabilidade por camadas foram incluídas no CODEC definido na norma MPEG-4. Esta norma inclui escalabilidade espacial e temporal através de um ambiente mais genérico e também acrescenta, no nível de objetos de vídeo, o conceito de Escalabilidade por Granularidade Fina (*Fine Grain Scalability*) (FGS), que é um tipo de escalabilidade SNR. Em seu perfil mais simples, o MPEG-4 é equivalente ao CODEC definido pela H.263, sem qualquer suporte à escalabilidade, embora extensões do H.263 também definam ferramentas de escalabilidade espacial, temporal e SNR.

Em qualquer dessas normas de escalabilidade de vídeo, no entanto, duas características sempre coexistem e devem ser ponderadas numa relação de compromisso. Por um lado, as ferramentas de codificação por camadas mostram-se muito ineficientes, principalmente por causa da estrutura recursiva do laço de predição inerente à combinação da compensação de movimento com a codificação por transformada em bloco dos codificadores híbridos. Por outro, como forma de melhorar a eficiência de codificação, pode-se utilizar informação das camadas superiores para a codificação de camadas inferiores, o que, por sua vez, causa o efeito *drift*, que é um erro acumulativo na decodificação das camadas intermediárias por falta de informação necessária das camadas suprimidas, no caso de recepção parcial.

Trabalhos como o de Wu et al. (2002) e os citados por Ohm (2005) foram realizados a fim de evoluir as técnicas de codificação escalável por camadas considerando o compromisso entre eficiência e *drift*. De forma geral, pode-se dizer que, idealmente, a predição de uma determinada camada não deveria utilizar informação de camadas superiores, pois assim evitaria estimativas diferentes entre o codificador e o decodificador, que geram os efeitos de propagação de erro. A predição que não utiliza a informação das camadas superiores, no entanto, sempre será pior do que poderia ser se utilizasse. No extremo deste segundo caso, ou seja, se toda a informação de enriquecimento fosse usada nas ferramentas de predição, a qualidade de reconstrução das camadas superiores de enriquecimento seria próxima, em desempenho, da qualidade de codificação em uma única camada. No entanto, a qualidade de reconstrução da camada base neste caso, e das eventuais camadas intermediárias, poderia sofrer de forma dramática com problemas de propagação de erro.

Por esses motivos, acreditava-se por volta de 2005 que a codificação por *wavelet*, intrinsecamente escalável e sem problema de *drift*, seria a solução que naturalmente substituiria a codificação híbrida tradicional. Essa constatação também foi relatada no trabalho de Ohm (2005), que apresenta o panorama da escalabilidade à época, os avanços nas ferramentas para os codificadores híbridos e a codificação por Filtragem Temporal com Compensação de Movimento (*Motion-Compensated Temporal Filtering*) (MCTF), classificada

pelo autor como desenvolvimento mais promissor na área, pois estabelecia as bases plenas de uma transformada *wavelet* tridimensional integrada à compensação de movimento.

Foi neste contexto que se iniciaram os trabalhos de especificação da extensão de escalabilidade do H.264/AVC, que será apresentada ainda neste capítulo. Como será visto, apesar do cenário desfavorável às ferramentas de codificação híbrida, ganhos expressivos foram obtidos com técnicas novas que permitiram melhor controle do compromisso entre eficiência e *drift*. Dessa forma, o custo computacional adicional das ferramentas baseadas em *wavelet* não justificou os benefícios dessas ferramentas durante o processo de especificação da extensão da norma e, novamente, a transformada *wavelet* ficou de fora. Por esse motivo, dá-se maior enfoque aos codificadores híbridos tradicionais neste capítulo, e os próximos itens, que tratam os tipos de escalabilidade, continuarão fazendo referências a esses CODECs.

2.4.1 Escalabilidade Espacial

Na escalabilidade espacial, a camada base possui uma versão de resolução reduzida de cada imagem codificada. Cada camada de enriquecimento representa um aumento de resolução em relação à camada anterior.

Para se codificar uma seqüência de vídeo com escalabilidade espacial em duas camadas, de forma simplista, são necessários os seguintes passos: primeiro é realizada uma subamostragem horizontal e vertical de cada quadro de vídeo de entrada para a codificação usual da camada base. Em seguida, a camada base é decodificada e tem cada quadro interpolado para se obter novamente a resolução original da seqüência. Cada quadro obtido por interpolação é subtraído então do quadro original e a diferença (resíduo) é então codificada para formar a camada de enriquecimento.

Esse mecanismo de codificação é denominado codificação espacial de laço único (apresenta um único laço de predição de movimento), e sua vantagem está na simplicidade. No entanto, como nesse mecanismo a informação da camada de enriquecimento não inclui laço de predição de movimento, a eficiência da codificação é comprometida, pois a informação de outros quadros (passados e futuros) não é considerada para eliminar redundâncias do quadro atual.

A escalabilidade espacial definida no MPEG-2 e no MPEG-4 já é mais sofisticada, usa dois laços de predição, um na camada base e outro na camada de enriquecimento. O decodificador escalável MPEG-2 usa como predição uma combinação ponderada de um

quadro interpolado reconstruído da camada base com o último quadro reconstruído da camada de enriquecimento. O decodificador MPEG-4 escalável, por sua vez, permite uma predição *bi-direcional*, utilizando um quadro reconstruído da camada base como *quadro passado* e o último quadro reconstruído da camada de enriquecimento como *quadro futuro* (CHOUPANI; WONG; TOLUN, 2007).

2.4.2 Escalabilidade Temporal

A camada base de uma seqüência codificada por escalabilidade temporal é formada por uma baixa taxa de quadros e a camada de enriquecimento temporal é formada por quadros que são intercalados aos quadros da camada base aumentando assim a taxa de quadros da seqüência final. Ambas as camadas possuem quadros codificados com a mesma resolução espacial.

A eficiência de codificação da escalabilidade temporal é alta, muito próxima da codificação sem escalabilidade, ou até maior como em casos específicos com o H.264/AVC. Isso se deve ao fato de a escalabilidade temporal ser implementada, normalmente, apenas como um conjunto de regras sobre a ordem de codificação dos quadros e, conseqüentemente, das possíveis referências entre quadros. Parte significativa da perda de eficiência, quando ocorre, é decorrente do maior intervalo de tempo entre quadros na camada base, o que diminui a correlação temporal entre quadros consecutivos. Os quadros da camada base são, geralmente, codificados como quadros tipo I ou P, enquanto que os quadros da camada de enriquecimento podem ser tipo I, P ou B.

2.4.3 Escalabilidade por Qualidade

Escalabilidade por qualidade ou por taxa de bits ou ainda SNR é a técnica de codificação de uma seqüência de vídeo em camadas com mesma resolução espacial e mesma taxa de quadros, porém com diferentes precisões na etapa de quantização.

No decodificador definido pelo MPEG-2, o fluxo de bits é decodificado primeiro pelo Decodificador de Comprimento Variado (*Variable Length Decoder*) (VLD) da camada base (CHOUPANI; WONG; TOLUN, 2007). Em seguida, o quantizador inverso dessa mesma camada reconstrói, a partir dos resultados anteriores, os coeficientes da DCT. Em paralelo, o fluxo de bits da camada de enriquecimento é processado por mecanismo idêntico ao descrito para a camada base, reconstruindo o resíduo de enriquecimento para os coeficientes DCT. Pela adição dos dois resultados anteriores são obtidos coeficientes de maior

precisão. Estes coeficientes são passados para a unidade de Transformação Inversa do Cosseno (*Inverse Discrete Cosine Transform*) (iDCT) para produzir blocos de pixels no domínio da imagem reconstruída. Em caso de quadros do tipo P, o resultado da transformação inversa constitui resíduos que devem ser adicionados ao bloco de compensação de movimento de um quadro anterior.

Na codificação por escalabilidade SNR do MPEG-2, a informação da camada de enriquecimento pode ser usada no laço de predição de movimento. Os resultados dessa codificação evidenciam a relação de compromisso entre eficiência e *drift* citada anteriormente, de onde conclui-se que ou a camada base tem um desempenho ruim para garantir bom desempenho da camada de enriquecimento, ou o contrário (CHOUPANI; WONG; TOLUN, 2007).

2.4.4 Escalabilidade por Controle Granular de Taxa

O controle granular da taxa de codificação de um fluxo de bits escalável (codificação FGS) foi incorporado ao MPEG-4 pela crescente demanda de aplicações de vídeo por rede. O conceito fundamental do FGS é proporcionar controle altamente flexível sobre a taxa de bits do fluxo de vídeo, codificando-se a camada de enriquecimento de tal forma que os coeficientes resultantes da DCT são representados em bits que são reagrupados em planos, dos bits mais significativos para os menos significativos. A quantidade de planos que são transmitidos pode ser truncada durante a codificação ou depois dela, reduzindo a taxa de bits e a qualidade do vídeo conforme necessário.

FGS é fundamentalmente diferente de qualquer outra codificação sem escalabilidade ou com escalabilidade por camadas, pois permite que a codificação seja feita sem que a taxa de transmissão do fluxo de bits final seja conhecida a priori (LI, 1999).

Do lado do codificador, resumidamente, a ferramenta FGS do MPEG-4 opera obtendo o resíduo entre os coeficientes DCT originais e os recalculados a partir da decodificação do bloco equivalente de referência (normalmente um bloco da camada base). Em seguida, codificam-se os valores absolutos dos resíduos em planos de bits e por fim codificam-se seletivamente os sinais dos resíduos (CHOUPANI; WONG; TOLUN, 2007).

2.5 Norma de Codificação H.264/AVC

Com as informações essenciais sobre escalabilidade apresentadas, será passada uma visão geral sobre as características da norma H.264/AVC para posterior detalhamento de sua

extensão de escalabilidade.

A norma H.264/AVC, disponível como ITU-T H.264 ou MPEG-4 Parte 10 AVC, é ainda considerada estado-da-arte em codificação de vídeo digital e está sendo desenvolvida em quatro fases pelo *Joint Video Team* (JVT).

A primeira fase, concluída em 2003, consistiu na padronização dos perfis básicos de codificação. Esta fase foi iniciada em 1998, quando o VCEG publicou uma chamada para propostas no âmbito do projeto até então denominado H.26L. Em Outubro de 1999 foi publicado o primeiro rascunho da norma e em Dezembro de 2001, o grupo VCEG uniu-se finalmente ao MPEG. Formou-se assim o JVT, que finalmente publicou a primeira versão final da norma em Março de 2003.

A segunda fase, concluída parcialmente em 2005, consistiu na padronização de novos perfis para extensão de fidelidade. Em 2007, ainda foram publicados novos perfis com ferramentas de codificação para suporte à qualidade profissional.

A terceira fase, que engloba os trabalhos relacionados às ferramentas de escalabilidade, foi parcialmente concluída com a publicação da primeira extensão em 2007. Novos tipos de extensões de escalabilidade continuam sendo investigados e devem gerar novas versões da norma.

A quarta fase, ainda em execução, consiste na padronização da extensão para suporte à codificação multicâmera. Conforme planejamento do grupo, esta última fase deve ser concluída, pelo menos parcialmente, em 2009.

Estas fases, no entanto, não encerram iniciativas de avaliação de outras ferramentas ou a busca por conceitos de codificação inovadores. É possível a criação tanto de uma nova norma de codificação de vídeo, que seria possivelmente nomeada ITU-T H.265 como já cogitado, quanto de novas extensões para o próprio H.264/AVC, como tem mostrado a tendência.

2.5.1 Principais Características e Funcionalidades

No estágio atual da norma, com metodologias e características como uma estimação de movimento com precisão de 1/4 de pixel, blocos de tamanho variável, quadros de referência múltiplos e codificação de entropia baseada em contexto com códigos variáveis, o H.264/AVC proporciona, em média, um aumento de mais de duas vezes na eficiência de codificação em relação ao padrão mais difundido, o MPEG-2, em termos de relação entre qualidade e taxa.

De uma forma geral, a arquitetura do codificador de vídeo H.264/AVC, apresentado na Figura 2.6, é semelhante à do codificador da norma H.262/MPEG-2: está baseada nas previsões intraquadro e entre quadros, numa transformada estilo DCT, em quantização dos coeficientes da DCT e em métodos de codificação por entropia.

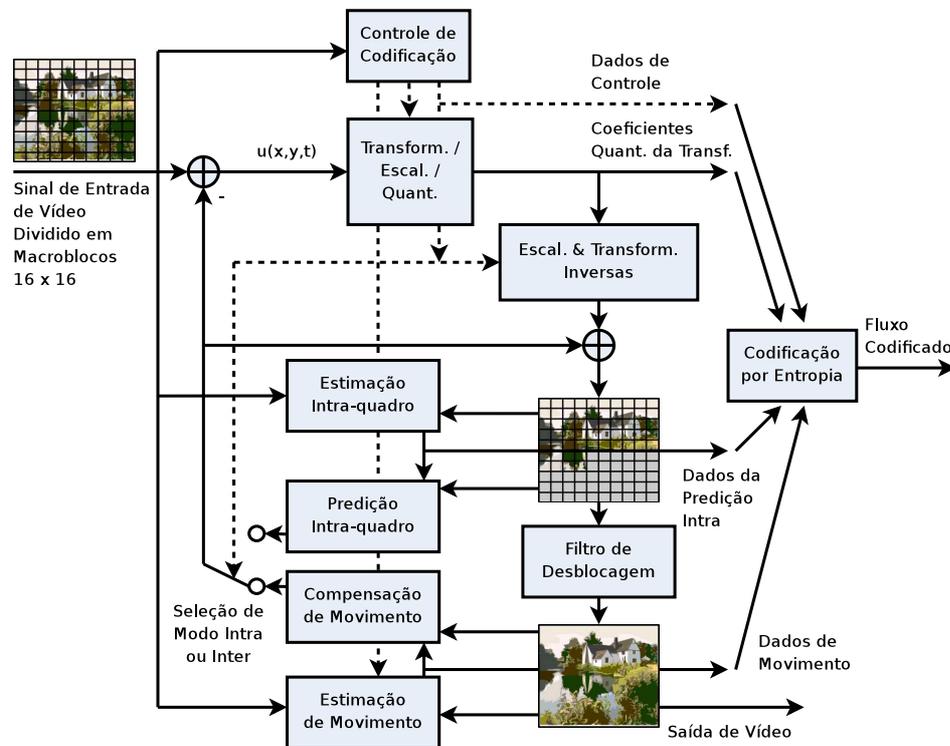


Figura 2.6: Arquitetura do codificador de vídeo H.264/AVC

Além destes blocos, que apesar de já serem conhecidos, apresentam características novas em relação a normas anteriores, outros blocos são exclusivos do H.264/AVC. Destaca-se o filtro de desblocagem, que anteriormente era utilizado como ferramenta opcional de pós-processamento no decodificador e passou a fazer parte do laço principal de codificação.

O ganho de eficiência de codificação do H.264/AVC não está especificamente em uma ou outra funcionalidade, ferramenta ou melhoria, mas sim, no conjunto. As funcionalidades mais importantes que contribuem para esse ganho expressivo de eficiência em suas ferramentas de predição e estimção de movimento são as seguintes, conforme resumido por Wiegand et al. (2003):

Compensação de movimento em blocos de tamanho variável: a norma permite maior flexibilidade na seleção de tamanho e forma dos blocos para compensação de movimento, permitindo o particionamento dos macroblocos de luma em subpartições tão pequenas quanto 4x4 pixels, como mostra a Figura 2.7.

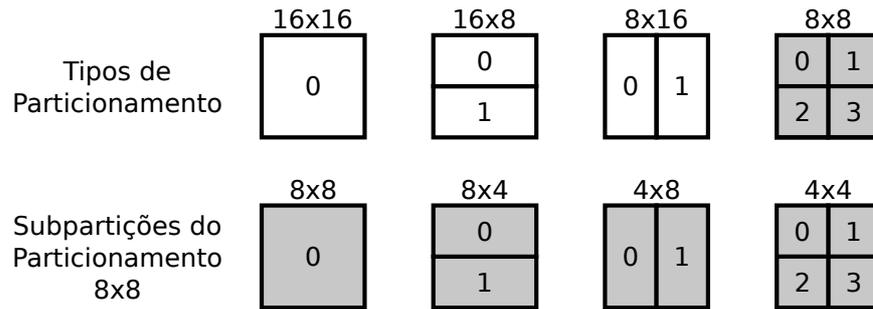


Figura 2.7: Tamanhos de blocos de predição de movimento no H.264/AVC

Compensação de movimento com precisão de 1/4: a maior parte das normas anteriores permitia precisão de no máximo meia amostra no vetor de movimento. Assim como primeiro visto no perfil avançado do MPEG-4, o H.264/AVC permite precisão de um quarto de amostra no vetor de movimento.

Vetor de movimento extrapolando os limites da figura: enquanto no MPEG-2 e seus predecessores o vetor de movimento estava restrito à área interna da figura de referência, assim como no H.263, esta norma também inclui a técnica de extrapolação dos limites da figura de referência.

Compensação de movimento com múltiplas figuras de referência: no MPEG-2 e seus predecessores, os índices das figuras referenciadas na compensação de movimento eram pré-definidos pela norma. O H.264/AVC estende a técnica já aprimorada no H.263++ para permitir flexibilidade no momento de apontar o índice da figura de referência na composição de figuras tipo P e B.

Desacoplamento entre ordem de referência e ordem de apresentação: em normas anteriores havia dependência estrita entre a ordem de referência das figuras para compensação de movimento e a ordem para apresentação. No H.264/AVC não há imposição sobre a seqüência de referência de figuras para compensação de movimento ou apresentação, havendo limitação neste mecanismo apenas pela quantidade de memória especificada para garantir capacidade de decodificação.

Uso irrestrito de figuras como referência: em normas anteriores, uma figura do tipo B não podia ser usada como referência na compensação de movimento. Na nova norma, essa restrição não existe. Essa liberdade foi fundamental para a codificação hierárquica por quadros B, que é a base para a codificação escalável do H.264/AVC, como será visto.

Predição ponderada: uma inovação no H.264/AVC permite o sinal da predição de com-

pensação de movimento ser ponderado e deslocado por quantidades especificadas pelo codificador. Isso pode aumentar significativamente a eficiência de codificação em cenas com esmaecimento, além de favorecer outras situações também.

Inferência de movimento melhorada: em normas anteriores, uma área “pulada” na codificação preditiva não podia apresentar movimento no conteúdo de cena. Isso era uma restrição para cenas com movimentação global. Por isso, o H.264/AVC infere o movimento em áreas “puladas”. Para áreas do tipo B, a norma ainda especifica um método conhecido como compensação de movimento “direta”, que melhora ainda mais o mesmo método já disponível no H.263 e no MPEG-4 Visual.

Predição espacial direcional para codificação intra: um novo método de extrapolação das bordas de blocos já decodificados de figuras do tipo I é aplicado para prever regiões ainda não decodificadas. Isso aumenta significativamente a compressão de blocos tipo I, que são os mais custosos em termos de taxa. A Figura 2.8 apresenta os nove modos de predição intra de macroblocos subdivididos em blocos de 4x4 pixels. Além destes modos, ainda há mais quatro modos aplicáveis a macroblocos inteiros, ou seja, sem subdivisão.

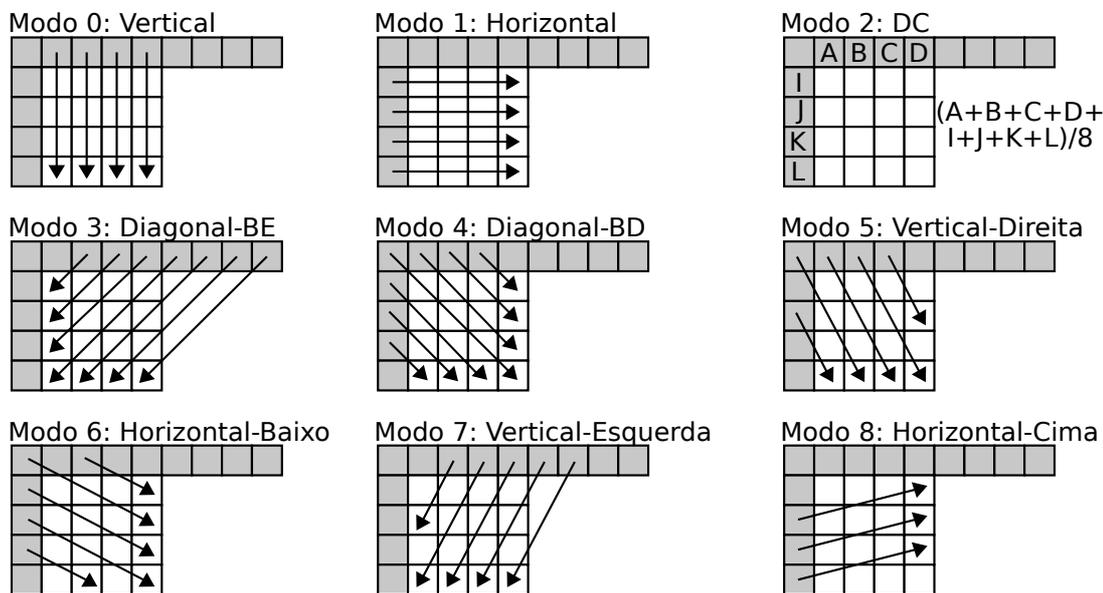


Figura 2.8: Modos de predição intra de macroblocos subdivididos em blocos 4x4 na norma H.264/AVC

Filtragem de desbloqueamento dentro do laço: conforme já mencionado, filtros de desbloqueamento já eram utilizados como ferramenta de pós-processamento para reduzir as marcas em fronteiras entre blocos da codificação. Com a passagem desta ferra-

menta para o codificador, há tanto uma diminuição dos efeitos de blocagem quanto um aumento de eficiência nas ferramentas de predição e compensação de movimento.

Além dos aprimoramentos nos métodos de predição, outras partes da norma H.264/AVC que também se beneficiaram com melhor eficiência de codificação, complementando o que Wiegand et al. (2003) apresenta, incluem:

Transformada em pequenos blocos: enquanto a maioria das normas anteriores realizava a transformada em blocos de tamanho 8x8, o H.264/AVC realiza em blocos 4x4.

Transformada em blocos hierárquicos: na maior parte dos casos, a nova transformada de blocos 4x4 é benéfica. No entanto, em algumas situações em que a correlação entre pixels é grande em regiões amplas da imagem, é possível realizar uma transformada hierárquica que estende o tamanho dos blocos de croma e luma para 8x8 e 16x16 respectivamente, aproveitando melhor a correlação da imagem.

Transformada com palavras de menor comprimento: todas as normas anteriores necessitavam de processamento mais complexo no codificador e no decodificador para a computação das transformadas. Enquanto os projetos anteriores necessitavam 32 bits de processamento, o H.264/AVC necessita apenas aritmética com 16 bits.

Transformada inversa com cálculo exato: as transformadas usadas em codificação de vídeo eram geralmente especificadas somente dentro de uma faixa de tolerância, pois era impossível a obtenção de um resultado exato em função de arredondamentos envolvidos nos cálculos digitais. Com isso, cada decodificador podia produzir resultados ligeiramente diferentes, causando propagação de erro na decodificação. O H.264/AVC defini uma transformação exata que garante o mesmo resultado de decodificação em qualquer decodificador.

Codificação aritmética de entropia: além de suporte padrão à codificação por código de comprimento variado, o H.264/AVC inclui suporte a outro método avançado conhecido como codificação aritmética.

Codificação de entropia adaptada a contexto: tanto Codificação por Comprimento Variado Adaptada ao Contexto (*Context-Adaptive Variable Length Code*) (CAVLC) quanto Codificação Binária Aritmética Adaptada ao Contexto (*Context-Adaptive Binary Arithmetic Coding*) (CABAC) no H.264/AVC melhoraram a eficiência de

codificação em relação às normas anteriores por considerar a informação do contexto da imagem na codificação.

Codificação adaptativa para entrelaçamento: duas ferramentas distintas de codificação de vídeo entrelaçado estão disponíveis no H.264/AVC, aprimorando as técnicas utilizadas em normas anteriores: *Picture Adaptive Frame Field* (PAFF) e *Macroblock Adaptive Frame Field* (MBAFF). A primeira permite que cada quadro seja codificado em uma única figura, representando o quadro todo, ou em duas figuras, sendo uma para cada quadro. A segunda ferramenta permite que essa decisão seja realizada a cada par vertical de macroblocos de um quadro, sendo mais adequada para cenas com movimentos localizados.

Para atender aos requisitos de flexibilidade e possibilidade de customização para a ampla faixa de aplicações a que se dispunha atender o H.264/AVC, foram padronizadas duas camadas na norma: Camada de Codificação de Vídeo (*Video Coding Layer*) (VCL), projetada para representar de forma eficiente o conteúdo de vídeo e Camada de Abstração de Rede (*Network Abstraction Layer*) (NAL), que formata o conteúdo VCL e provê informações de cabeçalho de maneira apropriada para uma variedade de camadas de transporte e mídias de armazenamento.

Robustez a perdas e erros de dados e flexibilidade para operação em uma variedade de ambientes de redes também são contemplados por alguns outros aspectos de projeto do H.264/AVC, que novamente segundo Wiegand et al. (2003) incluem:

Estrutura de configuração de parâmetros: algumas informações essenciais sobre o vídeo codificado foram tratadas em estruturas especiais, evitando o que ocorria em normas anteriores, que possibilitavam impacto muito negativo pela possível perda de bits chaves do fluxo.

FMO e ASO: no H.264/AVC são providas formas mais flexíveis de agrupamento de macroblocos em fatias (grupos de macroblocos de uma imagem) e também formas mais flexíveis de ordenação das fatias de uma imagem, permitindo que elas sejam decodificadas independentemente.

Figuras redundantes: para aprimorar a robustez à perda de dados o H.264/AVC permite o envio redundante de regiões de uma figura.

Partição de dados (*data partitioning*): como algumas informações codificadas de uma imagem podem ser mais relevantes que outras, o H.264/AVC permite que

os elementos sintáticos de cada fatia de macroblocos sejam agrupados em até três partições para transmissão.

SP/SI: figuras de sincronismo (SP) e de chaveamento (SI) foram criadas no H.264/AVC para permitir a troca de decodificador entre representações de vídeo usando taxas distintas, recuperação de perda de dados e erros e também modos avançados de controle como avanço e retrocesso rápido de vídeo.

2.5.2 Perfis (*Profiles*)

Como o padrão H.264/AVC tem um conjunto amplo de ferramentas e opções de codificação, nem todas as ferramentas são necessárias para todas as aplicações. Por exemplo, a ferramenta de correção de erros não é importante em aplicações que executam sobre uma rede robusta. Se todos os codificadores e decodificadores suportassem todas as ferramentas especificadas, eles teriam complexidade e custos desnecessários. Para contornar esse problema, subconjuntos de ferramentas foram definidos no H.264/AVC e denominados *Perfis*, como normas anteriores já haviam feito.

Em sua primeira versão, o H.264/AVC definiu três perfis básicos:

Perfil BASELINE: inclui codificação intra e inter quadros, permitindo fatias do tipo I e P, e codificação por entropia com CAVLC.

Perfil MAIN: inclui suporte a vídeo entrelaçado, codificação entre quadros usando fatias do tipo B, codificação entre quadros usando predição ponderada e codificação por entropia usando CABAC.

Perfil EXTENDED: não suporta vídeo entrelaçado ou CABAC, mas inclui modos que habilitam um chaveamento eficiente entre quadros pré-codificados (fatias SI e SP) e maior robustez a erros através da partição de dados.

Aplicações potenciais para o perfil BASELINE são: vídeo-telefone, vídeo-conferência e comunicações *wireless*. Para o perfil MAIN, as aplicações potenciais são: radiodifusão de TV e armazenamento de vídeo. E para o perfil EXTENDED, a principal aplicação é o *streaming* de mídia. Entretanto, cada perfil é suficientemente flexível para suportar uma vasta gama de aplicações e estes exemplos não são restritivos.

Além dos perfis originais, com foco em qualidade para entretenimento, a Extensão de Faixa de Fidelidade (*Fidelity Range Extensions*) (FRExt) de 2003 foi publicada para

permitir codificação de vídeo em qualidades superiores, pois ainda que o perfil MAIN possibilite resultados de PSNR muito superiores ao MPEG-2, o ganho de qualidade objetiva não se reflete em ganhos proporcionais de qualidade subjetiva (CHEN et al., 2005). Em particular, o perfil MAIN não é satisfatório na preservação de texturas refinadas em material de alta resolução. A extensão FExt adicionou os seguintes recursos principais:

- Transformadas inteiras de blocos 8x8, mais adequadas para alta resolução.
- Matrizes para escala de quantização ponderada.
- Mais de 8 bits nas amostras de luma e croma para melhorar a precisão.
- Formatos de amostragem 4:2:2 e 4:4:4, com resoluções superiores de croma.
- Codificação sem perdas entre figuras.
- Suporte a espaços de cores adicionais.

Esses recursos foram agrupados em quatro perfis adicionais, denominados perfis HIGH, cujas principais características são:

Perfil HIGH (HiP): suporta vídeo com amostras no formato 4:2:0 e 8 bits de resolução, destinado para consumidores *high-end* e outras aplicações que necessitam alta resolução sem formatos estendidos de croma e sem melhoria na precisão de amostras.

Perfil HIGH 10 (Hi10P): suporta vídeo no formato 4:2:0 com até 10 bits por amostra.

Perfil HIGH 4:2:2 (Hi422P): suporta vídeo em formato até 4:2:2 com até 10 bits por amostra.

Perfil HIGH 4:4:4 (Hi444P): suporta vídeo em formato até 4:4:4 com até 12 bits por amostra e, adicionalmente, suporta de maneira eficiente a codificação de partes da imagem sem perdas e transformada inteira da imagem residual para codificação de vídeo RGB.

Estes quatro perfis suportam ainda todas as ferramentas do perfil MAIN, tamanho adaptável de blocos e matriz de quantização perceptual escalável.

No entanto, a avaliação destes perfis pela indústria indicou que o Hi444P deveria ter sido projetado de maneira diferente, o que fez com que esse perfil fosse substituído e retirado da norma em nova extensão publicada em Abril de 2007. Esta extensão adicionou,

além do perfil Hi444PP que substituiu o Hi444P, mais quatro novos perfis de qualidade profissional que são, na verdade, subconjuntos de perfis já existentes na norma:

Perfil HIGH 4:4:4 Preditivo (Hi444PP): construído com base no Hi422P, suporta até o formato 4:4:4 de amostragem de cor, até 14 bits por amostra, codificação eficiente de regiões sem perdas e ainda codificação para cada figura de três planos independente de cor.

Perfil HIGH 10 Intra: perfil Hi10P restrito à codificação intra.

Perfil HIGH 4:2:2 Intra: perfil Hi422P restrito à codificação intra.

Perfil HIGH 4:4:4 Intra: perfil Hi444PP restrito à codificação intra.

Perfil CAVLC 4:4:4 Intra: perfil Hi444PP restrito à codificação intra e ao CAVLC.

Tabela 2.2: Ferramentas dos perfis básicos e da extensão FRExt do H264/AVC

Ferramentas	BL	EXT	MAIN	HiP	Hi10P	Hi422P	Hi444PP
Fatias tipo I e P	X	X	X	X	X	X	X
Fatias tipo B		X	X	X	X	X	X
Fatias tipo SI e SP		X					
Múltiplos quadros de ref	X	X	X	X	X	X	X
Filtro de desblocagem	X	X	X	X	X	X	X
CAVLC	X	X	X	X	X	X	X
CABAC			X	X	X	X	X
FMO e ASO	X	X					
Fatias redundantes	X	X					
Partição de dados		X					
PicAFF e MBAFF		X	X	X	X	X	X
Formato de croma 4:2:0	X	X	X	X	X	X	X
Formato monocromático				X	X	X	X
Formato de croma 4:2:2						X	X
Formato de croma 4:4:4							X
Amostras de 8 bits	X	X	X	X	X	X	X
Amostras de 9 e 10 bits					X	X	X
Amostras de 11 e 14 bits							X
Transformadas 4x4 e 8x8				X	X	X	X
Matrizes de escala de quant.				X	X	X	X
Quant. indep. Cb/Cr				X	X	X	X
Codif. separada de cores							X
Codif. preditiva sem perdas							X

A Tabela 2.2 resume as ferramentas suportadas em cada perfil do conjunto básico e da extensão FRExt, já considerando o perfil Hi444PP no lugar do Hi444P.

2.5.3 Níveis (*Levels*)

Enquanto os perfis do H.264/AVC definem o conjunto de ferramentas que podem ser usadas na codificação de uma seqüência de vídeo, os níveis definem restrições de parâmetros que determinam os limites de desempenho das ferramentas. Os níveis são importantes, acima de tudo, para definição da capacidade de processamento e de memória de acordo com o estritamente necessário na implementação de codificadores e decodificadores em função da aplicação. O tamanho das imagens, a taxa de quadros e a quantidade de imagens de referência são variáveis diretamente relacionadas a estes parâmetros.

A Tabela 2.3 apresenta os 16 níveis definidos na norma H.264/AVC e os principais parâmetros associados.

Tabela 2.3: Níveis do H.264/AVC

Níveis	Tamanho típico da imagem	Taxa de quadros típica	Taxa máxima (VCL) para perfis básicos	Número máximo de quadros de referência ^a
1	QCIF	15	64 kbps	4
1b	QCIF	15	128 kbps	4
1.1	CIF ou QCIF	7.5 (CIF) / 30 (QCIF)	192 kbps	2 (CIF) / 9 (QCIF)
1.2	CIF	15	384 kbps	6
1.3	CIF	30	768 kbps	6
2	CIF	30	2 Mbps	6
2.1	HHR (480i ou 576i)	30 / 25	4 Mbps	6
2.2	SD	15	4 Mbps	5
3	SD	30 / 25	10 Mbps	5
3.1	1280x720p	30	14 Mbps	5
3.2	1280x720p	60	20 Mbps	4
4	HD (720p or 1080i)	60p / 30i	20 Mbps	4
4.1	HD (720p or 1080i)	60p / 30i	50 Mbps	4
4.2	1920x1080p	60p	50 Mbps	4
5	2kx1k	72	135 Mbps	5
5.1	2kx1k ou 4kx2k	120 / 30	240 Mbps	5

^aConsiderando tamanho típico de imagens para o nível correspondente.

Uma vez que os perfis acrescentados à norma pela extensão FRExt são especificados para aplicações de alta demanda por qualidade, a máxima taxa de bits por nível, nestes casos, é incrementada por um fator multiplicativo. Além disso, algumas limitações adicionais são especificadas na norma para cada perfil. Nos perfis MAIN e HIGH, por exemplo, as ferramentas de entrelaçamento deixam de ser suportadas a partir do nível 4.2 (inclusive).

2.6 A Extensão de Escalabilidade no H.264/AVC

Depois do grande sucesso nas duas primeiras fases de padronização do H.264/AVC, que envolveram a criação dos perfis básicos e posteriormente das extensões de fidelidade, o JVT passou a trabalhar no desafio de construir, sobre as bases sólidas desse padrão, ferramentas para a codificação escalável de vídeo. O *Scalable Video Coding* (SVC), como foi denominado o resultado desse desafio, foi incorporado ao padrão H.264/AVC como uma extensão, complementando suas funcionalidades para endereçar a necessidade de aplicações por ferramentas mais flexíveis de codificação para uso em ambientes altamente heterogêneos e variantes no tempo (LUTHRA; SULLIVAN; WIEGAND, 2003).

A compreensão dos erros cometidos em iniciativas anteriores de padronização de ferramentas de escalabilidade fez com que os trabalhos do SVC gerassem resultados finalmente mais promissores em termos de expectativa de adoção em escala pela indústria. Anteriormente já haviam sido feitas ao menos outras quatro tentativas de padronização de escalabilidade em vídeo (uma para o MPEG-2, uma para o H.263 e outras duas para gerações do MPEG-4 Parte 2) e a principal lição aprendida com essas tentativas foi de que o projeto do SVC deveria ser mantido o mais simples possível (WIEGAND et al., 2007). Esse requisito foi assegurado por um rígido processo de avaliação das ferramentas inicialmente propostas, de forma que foram removidas aquelas que não se mostraram realmente necessárias, cujo mérito não foi comprovado ou que ainda não haviam atingido grau de maturidade satisfatório.

Por esses motivos, ainda que das 14 propostas submetidas à chamada inicial do grupo, 12 sugerissem o uso de transformadas por 3-D *wavelets*, depois de uma fase de seis meses de análise, essas propostas foram descartadas (SCHWARZ; MARPE; WIEGAND, 2007). O SVC baseou-se, dessa forma, na codificação por camadas conforme modelo inicial proposto em Schwarz et al. (2004), especificando ferramentas de escalabilidade que se integram à estrutura híbrida pura do H.264/AVC.

Além disso, assim como nos padrões precedentes de codificação de vídeo, no SVC foram especificados somente o formato da sintaxe do fluxo de bits e o processo de decodificação. Dessa forma, funcionalidades que puderam ser resolvidas sem prejuízo apenas com pré-processamento ou técnicas exclusivas de codificação, por exemplo, não foram incluídas no padrão. Em alguns casos também foi possível prover funcionalidades de escalabilidade através da especificação de informações suplementares ao invés de se modificar o núcleo principal da sintaxe do fluxo ou do processo de decodificação.

Os principais resultados obtidos com o direcionamento dado aos trabalhos do SVC podem ser resumidos como:

Excelente eficiência de codificação: em cenários de aplicações típicas, com codificadores suficientemente otimizados, a codificação escalada apresenta uma sobrecarga apenas de cerca de 10% na taxa de bits.

Mínima complexidade computacional: foi evitada a necessidade de recursos computacionais excessivos, principalmente com a introdução do método de decodificação por laço simples (*single-loop decoding*).

Máxima consistência de projeto: manteve-se máxima similaridade entre as novas ferramentas e as pertencentes ao H.264/AVC, incluindo a possibilidade de tradução de um fluxo em camadas para um fluxo tradicional H.264/AVC sem necessidade de um processo de decodificação completo.

No SVC, seguindo o requisito de simplicidade, as camadas de enriquecimento de vídeo podem ser dos três tipos fundamentais: temporal, espacial e de qualidade, conforme descrito nos próximos itens com base no artigo de Schwarz, Marpe e Wiegand (2007) e exemplos deste mesmo texto, que apresenta uma ótima visão geral sobre o tema.

2.6.1 Escalabilidade Temporal

Pode-se considerar que a especificação original do H.264/AVC já previa ferramentas para escalabilidade temporal, pois em codificadores híbridos, este tipo de escalabilidade pode ser obtido através de restrições sobre os valores de índices possíveis para os quadros de referência na predição de movimento. Nesse sentido, o controle de memória das figuras de referência no H.264/AVC proporciona um aumento significativo de flexibilidade, uma vez que permite dependências temporais arbitrárias entre figuras, conforme visto na Seção 2.5.1. Assim sendo, para o suporte de escalabilidade temporal com um número razoável de camadas, nenhuma alteração foi necessária no projeto original de codificação do H.264/AVC, somente no formato da sinalização para diferenciar os dados de cada camada.

A Figura 2.9 apresenta uma possível estrutura em camadas da codificação escalável temporal no SVC, conhecida como *estrutura de predição hierárquica*, pela relação hierarquizada entre os quadros da seqüência. Cada cor da seqüência representa um conjunto de figuras de uma camada distinta, sendo que as camadas de enriquecimento (camadas T_x

com $x > 0$) são tipicamente compostas por figuras do tipo B, codificadas com referência a uma figura anterior e outra posterior, ambas de uma camada inferior a atual.

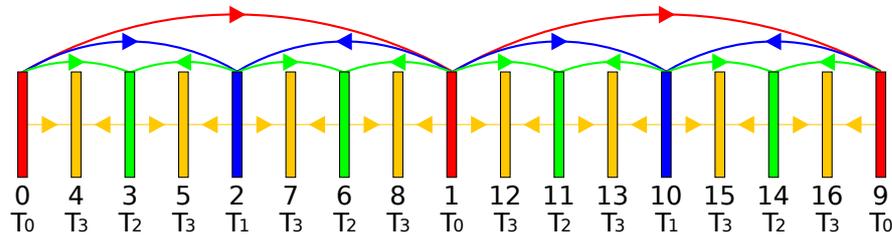


Figura 2.9: Escalabilidade temporal dicotômica no SVC

Embora este seja um caso particular, no qual as camadas de enriquecimento são compostas por intercalação progressiva de figuras entre figuras consecutivas de camadas inferiores (dicotomia), e no qual a estrutura de referências entre figuras é fixa, este é um caso de excelente eficiência de codificação (SCHWARZ; MARPE; WIEGAND, 2005).

A flexibilidade do H.264/AVC permite, no entanto, que a codificação hierárquica seja realizada também de forma não dicotômica, inserindo-se mais de uma figura entre duas consecutivas na composição de uma camada de enriquecimento qualquer, como mostra a Figura 2.10.

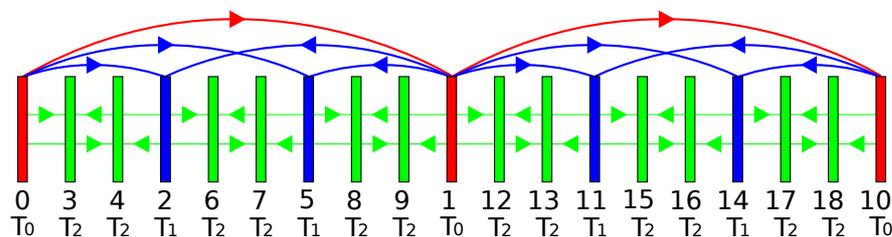


Figura 2.10: Escalabilidade temporal não dicotômica no SVC

Por fim, através ainda de restrições impostas sobre os índices das figuras que podem ser usadas como referência no mecanismo de predição é possível impor condições de atraso máximo de codificação e decodificação como mostra a Figura 2.11.

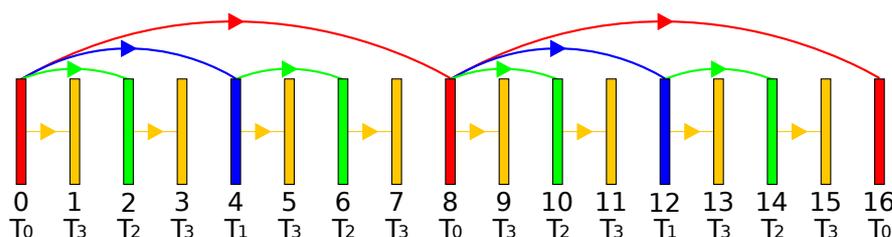


Figura 2.11: Escalabilidade temporal com atraso nulo no SVC

Neste esquema, o atraso é zero, pois uma figura depende apenas de figuras anteriores da seqüência, logicamente com compromisso de perda de eficiência de codificação.

De forma geral, a estrutura de codificação pode variar ao longo do tempo, para se adaptar às características locais da seqüência. Além disso, como a ordem de codificação de figuras de uma estrutura de predição hierárquica deve ser escolhida de forma que as figuras de referência sejam codificadas antes de poderem ser usadas pelo mecanismo de predição, existem inúmeras estratégias diferentes para se fazer essa escolha. As estratégias variam basicamente em termos de atraso na codificação e em requisitos de memória, influenciando a eficiência da codificação.

A eficiência da codificação hierárquica depende fortemente também da escolha dos parâmetros de quantização, que podem variar de camada para camada. Várias estratégias são propostas para essa escolha, como apontam Schwarz, Marpe e Wiegand (2007). Algumas estratégias são bastante complexas e exigem muitos cálculos, ao passo que outras, simplificadas, garantem ainda boa eficiência com menor custo computacional.

Intuitivamente, pode-se dizer que as figuras da camada base devem ser codificadas com a maior fidelidade, visto que estas serão utilizadas direta ou indiretamente na predição por compensação de movimento das figuras de todas as demais camadas. Como consequência dessa constatação, o parâmetro de quantização deve ser sucessivamente aumentado em relação ao anterior para cada camada codificada, pois cada vez menos figuras são influenciadas.

Na codificação hierárquica, sabe-se ainda que os quadros do tipo B apresentam melhor eficiência se gerados utilizando-se o recurso de compensação por soma ponderada dos blocos de referência.

Quando mais de duas camadas temporais são utilizadas, é também recomendado o uso do *modo espacial direto* do H.264/AVC, pois o modo temporal direto produz vetores inadequados para cerca de metade das figuras tipo B.

2.6.2 Escalabilidade Espacial (ou Escalabilidade de Resolução)

O suporte à escalabilidade espacial no SVC segue a solução convencional de codificação multicamada já utilizada no H.262/MPEG-2 Video, no H.263 e no MPEG-4 Visual.

No SVC, a resolução de uma camada não necessita ser um múltiplo inteiro da resolução da camada imediatamente anterior, não podendo apenas diminuir. As figuras de uma camada de enriquecimento podem representar ainda somente uma área retangu-

lar selecionada da camada inferior com maior ou igual resolução ou ainda podem conter áreas adicionais além das bordas das figuras de referência. Em cada camada podem ser empregadas predições por compensação de movimento ou intra da mesma forma que na codificação de uma camada simples. Para melhorar a eficiência da codificação em comparação à transmissão simultânea de duas resoluções diferentes codificadas em fluxos simples independentes, mecanismos adicionais de predição entre camadas (*inter-layer prediction*) foram incorporados à norma, como mostra a Figura 2.12.

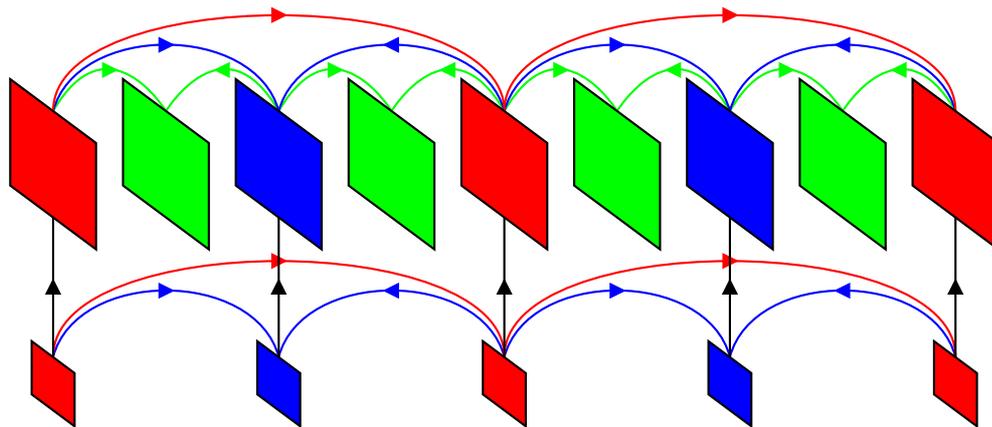


Figura 2.12: Escalabilidade espacial no SVC com predição entre camadas

Algumas restrições impostas pelo SVC amenizam os requisitos de memória e de complexidade computacional. Entre as restrições está o fato de que a mesma ordem de codificação de figuras deve ser usada em todas as camadas espaciais. O conjunto formado pelas figuras de cada camada de um mesmo instante de tempo representa uma *unidade de acesso*. Cada unidade de acesso deve ter suas figuras transmitidas em ordem, da camada base à última camada de enriquecimento. Como mostra a Figura 2.12, no entanto, não é necessário que as camadas inferiores apresentem figuras em todas as unidades de acesso, o que permite a combinação da escalabilidade espacial com a escalabilidade temporal.

Em normas anteriores ao SVC, o único método de predição entre camadas estava relacionado à disponibilização das amostras reconstruídas das camadas inferiores para as camadas de enriquecimento. Apesar destas amostras representarem toda a informação de uma determinada camada, elas não são necessariamente os dados mais adequados para a predição entre camadas. Como normalmente a predição entre camadas tem que competir com a predição temporal da camada em processo de codificação, normalmente o segundo mecanismo leva vantagem em termos de eficiência. Dessa forma, para melhorar a atratividade da predição entre camadas, foram adicionados ao SVC dois novos conceitos de predição: predição de modos do macrobloco e parâmetros de movimento associados e

predição de resíduo.

Estes dois mecanismos não excluem a predição usual entre figuras de uma mesma camada, pois estão restritos somente a figuras de uma mesma unidade de acesso. A predição usual, dessa forma, continua sendo a única opção de predição para uma figura cuja unidade de acesso não possua representantes nas camadas inferiores a sua. As ferramentas de predição entre camadas podem ser escolhidas até o nível de submacrobloco pelo codificador e um novo tipo de macrobloco pode ser sinalizado nas camadas de enriquecimento através do elemento sintático *base mode flag*.

Existem, contudo, restrições quanto aos tipos de macroblocos que podem ser usados como referência na codificação entre camadas para evitar complexidade computacional excessiva. Como resultado destas restrições, uma característica importante da codificação espacial do SVC é que cada camada de enriquecimento pode ser decodificada por um laço simples de compensação de movimento (SCHWARZ; MARPE; WIEGAND, 2007).

Em testes realizados com a escalabilidade espacial do SVC, deve-se notar que as estruturas hierárquicas são responsáveis por melhorar, além da eficiência de codificação de forma geral, a efetividade dos mecanismos de predição entre camadas.

2.6.3 Escalabilidade por Qualidade

A escalabilidade por qualidade no H.264/AVC também é chamada de Escalabilidade com Baixa Granularidade (*Coarse-Grain Quality Scalability*) (CGS) e trata-se de um caso particular da escalabilidade espacial, no qual as camadas de enriquecimento possuem a mesma resolução e taxa de quadros da camada base. No CGS os mesmos mecanismos de predição entre camadas da escalabilidade espacial são aplicados, mas sem a necessidade das operações de interpolação ou de desbloqueamento entre camadas. Além disso, as operações de predição intra entre camadas e de predição do resíduo são realizadas no domínio dos coeficientes das transformadas, reduzindo a complexidade das operações, sendo cada camada responsável pelo acréscimo de informação de textura à informação da camada anterior. O passo de quantização utilizado nos coeficientes de cada camada é gradativamente menor que o das camadas anteriores, aumentando o nível de detalhes a cada camada.

A estratificação por camadas do CGS, no entanto, permite que somente algumas taxas de bits pré-definidas sejam contempladas no fluxo. Sendo a quantidade de taxas permitidas igual à quantidade de camadas, o ganho de qualidade acontece em passos mais grosseiros, de onde vem o nome CGS. A troca entre camadas na decodificação só é possível

em pontos determinados, nos quais as unidades de acesso determinam que não haja mais referências a figuras de unidades anteriores. Quando a diferença de taxa relativa entre as camadas diminui, diminui também a eficiência da codificação.

Para aumentar a flexibilidade das taxas de bits na codificação e a robustez a erros, aumentando conseqüentemente a eficiência da codificação, foi incluída também no projeto do SVC uma variação do CGS, conhecida por Escalabilidade com Granularidade Média (*Medium-Grain Quality Scalability*) (MGS). As diferenças entre ambas são uma sinalização alto-nível modificada, que permite a troca entre taxas de bits em qualquer unidades de acesso do vídeo, e o conceito chamado de figura chave (*key picture*), que permite melhor ajuste do compromisso entre propagação de erro e eficiência de codificação nas estruturas hierárquicas.

Considerando também os novos conceitos apresentados pelo H.264/AVC, a Figura 2.13 mostra estruturas com características distintas de compromisso entre eficiência de codificação e propagação de erros.

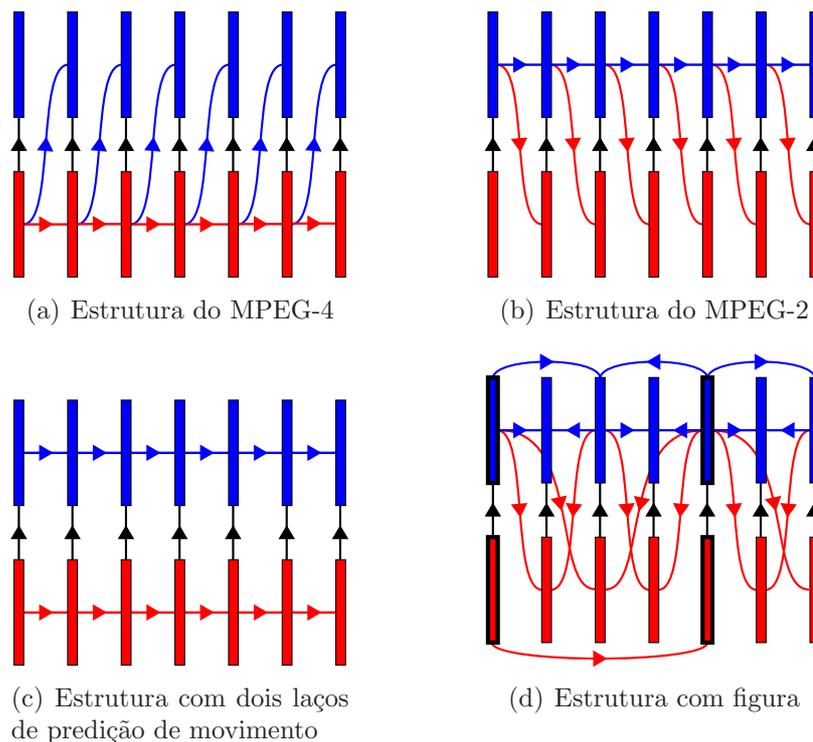


Figura 2.13: Tipos de estruturas de referência na escalabilidade por qualidade do SVC

Na Figura 2.13(a), observa-se a estrutura de codificação FGS do MPEG-4 Visual, escolhida de forma a anular completamente a propagação de erro por perda de pacotes da camada de enriquecimento. A contrapartida deste benefício, como já mencionado anteriormente, é a perda significativa de eficiência de codificação em comparação a uma

única camada.

Por outro lado, a Figura 2.13(b) apresenta o caso no extremo oposto, aplicado à escalabilidade por qualidade do H.262/MPEG-2 Video, na qual as figuras com maior nível de qualidade são usadas como referência para a predição de movimento. Neste caso tem-se alta eficiência de codificação nas camadas de enriquecimento, mas por outro lado alta suscetibilidade à propagação de erros por perda de pacotes.

Na seqüência, a Figura 2.13(c) apresenta um conceito alternativo de estrutura, na qual dois laços de predição podem ser empregados. Esta estrutura é similar à utilizada na escalabilidade espacial como especificado em H.262/MPEG-2 Video, H.263, e MPEG-4 Visual. Embora a perda de pacotes de enriquecimento não comprometa a qualidade da camada base, o erro provocado por esta perda propaga-se pelas demais figuras da própria camada de enriquecimento.

Por fim, a Figura 2.13(d) apresenta uma estrutura baseada no conceito de figura chave, introduzido no SVC para a codificação MGS. Bits especiais de sinalização indicam quais figuras (seja da camada base ou de enriquecimento) são utilizadas como referência na compensação de movimento, limitando os requisitos de memória. Adicionalmente, para limitar o custo extra de decodificação dessas figuras chaves, o SVC especifica que para elas os parâmetros de movimento não devem mudar entre a camada base e a de enriquecimento, o que permite um laço simples para a compensação de movimento dessas figuras. O conceito da figura chave pode ser combinado de forma muito eficiente com as estruturas de predição hierárquica. A combinação que está sendo apresentada possui figuras chaves nos instantes de tempo que configuram a resolução temporal mais grosseira da camada base, de forma que não haja propagação de erros na recomposição dessas figuras essenciais. Por outro lado, todas as demais figuras das camadas temporais superiores utilizam as figuras recuperadas das camadas de enriquecimento para a compensação de movimento, o que proporciona melhores resultados na eficiência da codificação. Como claramente as figuras chaves servem como pontos de re-sincronização entre codificador e decodificador, erros de perdas de pacotes propagam-se apenas por um período limitado. O compromisso entre eficiência de codificação e propagação de erro pode ser ajustado, dessa forma, em função da distância entre figuras chaves e da quantidade de estágios hierárquicos.

Todas as estruturas apresentadas na Figura 2.13 são casos particulares cobertos pela especificação do SVC.

2.6.4 Aplicação do SVC

De forma geral, a especificação do SVC é flexível para permitir diversas combinações entre os diversos tipos de escalabilidade vistos anteriormente, sendo que nenhum deles é obrigatório ou obriga o uso dos demais. Assim é possível escolher um ou mais tipos de escalabilidade em função da aplicação para maximizar a eficiência da codificação e minimizar os efeitos de propagação de erro.

Para permitir essa flexibilidade, a estrutura de sinalização no SVC divide as unidades de acesso e as informações associadas em camadas e sub-camadas. O primeiro nível de classificação da informação da unidade de acesso é o conjunto temporal a que ela pertence, identificado por um índice T, que varia de zero ao número máximo de camadas temporais do fluxo codificado menos um. Internamente, cada unidade de acesso é dividida em camadas espaciais, identificadas pelo grupo de resolução a que pertencem. O grupo por sua vez é identificado por um índice D, também no intervalo de zero ao número máximo de camadas distintas de resolução do fluxo de bits menos um. Por fim, cada camada de resolução pode ser subdividida em camadas de qualidade, identificadas por um índice Q, que varia de zero ao número máximo de camadas de qualidade para aquela determinada camada de resolução menos um. Com esta sinalização, a troca de resolução durante a decodificação entre camadas com Ds distintos só pode acontecer em figuras chaves (CGS), enquanto que a troca entre Qs distintos de uma mesma camada D pode ocorrer em qualquer unidade de acesso (MGS).

Além dos tipos já mencionados de escalabilidade, o SVC também oferece suporte à escalabilidade por codificação de Região de Interesse (*Region of Interest*) (ROI). Essa funcionalidade é realizada pela codificação diferenciada entre os macroblocos agrupados em fatias distintas na imagem. A principal limitação desse mecanismo é que as fatias de macroblocos devem respeitar alguns padrões pré-definidos.

Como um dos requisitos do SVC é a facilidade de extração de sub-fluxos de bits garantindo compatibilidade com sistemas anteriores do H.264/AVC, apesar de o cabeçalho da unidade NAL ter sido estendido de um para três bytes no SVC, cada fluxo de bits compatível com o SVC deve incluir um sub-fluxo básico composto por unidades não SVC, com cabeçalhos de um byte. Dessa forma, além de não impedir o funcionamento de equipamentos legados, os três bytes adicionais do SVC permitem fácil separação das camadas, que são identificadas nestes bytes adicionais pelos índices descritos anteriormente.

2.6.5 Perfis e Níveis do SVC

As ferramentas e estruturas de sinalização adicionadas com a extensão SVC da norma H.264/AVC foram organizadas em três novos perfis de escalabilidade, que correspondem a três perfis da versão anterior da norma, enquanto os níveis não foram alterados, permanecendo os mesmos:

Perfil BASELINE Escalável: essencialmente direcionado para aplicações de vídeo conferência e segurança, este perfil requer baixa complexidade de codificação. A relação entre as resoluções de camadas espaciais sucessivas está restrita a um e meio e dois e as ferramentas para conteúdo entrelaçado não estão presentes neste perfil.

Perfil HIGH Escalável: projetado para aplicações de radiodifusão, de transmissão por rede e de armazenamento, não há restrições quanto à relação entre resoluções das camadas neste perfil. As codificações por qualidade e temporal também são suportadas sem restrições.

Perfil HIGH INTRA Escalável: desenvolvido para contemplar principalmente aplicações profissionais, este perfil contém apenas figuras do tipo IDR, embora suporte o mesmo conjunto de ferramentas do perfil HIGH escalável.

2.7 Conclusão

Neste capítulo foi apresentada uma trajetória histórica das técnicas de codificação de vídeo com ênfase no estado-da-arte da codificação H.264/AVC, que inclui escalabilidade de vídeo.

Inicialmente, foram abordadas questões relacionadas com a análise de qualidade de vídeo e com a evolução dos codificadores, especificamente detalhando os codificadores híbridos ISO/ITU e citando outros codificadores, como os baseados em *wavelets*.

Este capítulo abordou de forma mais aprofundada a questão da escalabilidade de vídeo, discutindo características dessa técnica com base nos CODECs das famílias MPEG/H.26x. Os vários tipos de escalabilidade foram apresentados: temporal, espacial, por qualidade e granular.

Finalmente o capítulo apresentou a norma de codificação de vídeo H.264/AVC e sua extensão, revisitando e discutindo as questões da escalabilidade temporal, espacial e por qualidade no âmbito desta norma.

As evidências teóricas e o esforço de sistematização a cerca das questões apresentadas neste capítulo levaram o autor a concluir que o SVC é de potencial interesse para o SBTVD, que será apresentado no Capítulo 3.

3 Sistema Brasileiro de TV Digital

O Sistema Brasileiro de TV Digital (SBTVD), lançado em 02 de Dezembro de 2007 em São Paulo, aos poucos vai sendo disseminado pelo território nacional. Esperado por muitos já há algum tempo, principalmente pelos radiodifusores, que vêem na digitalização uma grande oportunidade de manutenção e expansão de seus negócios, o SBTVD possui um plano de implantação que prevê a desativação completa das transmissões analógicas, iniciadas há mais de 50 anos, até o final de 2016.

O esforço de transição da TV analógica para a TV digital, para se chegar ao estágio atual de implantação, foi e ainda é complexo no seu domínio social, técnico, político e econômico. Desde os primeiros grupos de trabalho constituídos ao longo da década de 90 para estudos em TV digital até os trabalhos recentes de normalização do sistema, muitos foram os acontecimentos. A televisão analógica está presente em aproximadamente 95% dos lares brasileiros, sendo ela responsável pela absorção de 61% (CONSELHO NACIONAL DE ENGENHARIA, 2007) da verba publicitária nacional, em seu modelo de gestão privada e transmissão aberta e gratuita. Era de se esperar que a sua digitalização despertasse tantos interesses e discussões.

De estudos comparativos entre os principais sistemas internacionais existentes (americano, europeu e japonês) a uma iniciativa nacionalista de definição de um sistema próprio com foco na inclusão social e na retomada do desenvolvimento tecnológico no país após muitos anos de pouco incentivo, a história da TV digital no Brasil reuniu os mais distintos personagens, incluindo governantes, radiodifusores, produtores de conteúdo, acadêmicos, desenvolvedores, fabricantes.

Com requisitos contraditórios num primeiro momento, objetivando baixo custo e ao mesmo tempo uso de tecnologia de ponta e inovação, o SBTVD foi proposto como mais uma alternativa frente aos concorrentes internacionais e não como uma decisão certa. Depois de muito trabalho, de pressões políticas e de muita turbulência pelas disputas de interesses, o SBTVD mostrou-se viável e foi finalmente escolhido para a digitalização da TV terrestre no Brasil.

3.1 TV no Brasil: das Transmissões Analógicas ao SBTVD

Em 2007, segundo dados do censo brasileiro (página de internet)¹, havia 54 milhões de aparelhos televisores no Brasil, distribuídos por aproximadamente 95% dos lares. Por sua abrangência expressiva, a televisão pode ser considerada hoje o principal meio de comunicação do país.

A história da TV analógica começou através da iniciativa do jornalista Francisco de Assis Chateaubriand Bandeira de Melo, que promoveu a primeira transmissão terrestre no Brasil no dia 03 de Abril de 1950 para alguns aparelhos instalados no saguão dos Diários Associados, dos quais era dono. Em 18 de Setembro do mesmo ano, Chateaubriand inaugurou a TV Tupi em São Paulo no canal 3, espalhando a modesta quantidade de 200 aparelhos de TV pela cidade devido ao alto preço e necessidade de importação.

Aos poucos o custo do aparelho televisor caiu, tornando-o mais popular. Na década de 70 o sistema tornou-se colorido mediante a criação do PAL-M, que foi desenvolvido no Brasil pela união do padrão de sinalização de cores europeu denominado Linhas com Alternância de Fase (*Phase Alternation Line-rate*) (PAL) com o padrão M de varredura (29,97 quadros por segundo), alinhado com a política protecionista do governo militar.

No final da década de 80 e início da década de 90, seguindo os movimentos mundiais já no sentido de digitalizar os sistemas de vídeo, o Brasil também deu seus primeiros passos como mostra a Figura 3.1, que apresenta os principais acontecimentos na história da digitalização da TV aberta no país.

A história da digitalização pode ser dividida em três fases principais: início dos trabalhos e primeiros testes; financiamento para pesquisa e tomada de decisão; implantação da TV digital terrestre.

3.1.1 Primeiros Trabalhos e Testes com Sistemas Digitais de Vídeo

Após algumas primeiras iniciativas de criação de núcleos de vídeo digital e de grupos técnicos para tratar o assunto de TV digital, a primeira fase da implantação do sistema de TV digital terrestre no Brasil, bem descrita por Junior (2008), incluiu os testes de campo realizados em São Paulo com o intuito de comparar o desempenho das versões terrestres

¹IBGE. Apresenta dados do *Instituto Brasileiro de Geografia e Estatísticas*. Disponível em: www.ibge.gov.br. Acesso em: 20 Ago. 2008.



Figura 3.1: Processo de definição da TV digital no Brasil

dos três sistemas disponíveis naquele momento: o sistema americano *Advanced Television Systems Committee* (ATSC), que depois de um longo desenvolvimento entrou em operação nos Estados Unidos em 1998; o sistema europeu *Digital Video Broadcasting* (DVB), que entrou em operação na Inglaterra também no ano de 1998; e o sistema japonês *Integrated Services Digital Broadcasting* (ISDB), que em fase de conclusão, entrou em operação somente em 2000.

A montagem do laboratório de testes e a execução dos mesmos foram realizadas pela Universidade Presbiteriana Mackenzie em São Paulo em cooperação com o Centro de Pesquisa e Desenvolvimento (CPqD) de Campinas. As principais dificuldades identificadas para o funcionamento de um sistema digital terrestre no Brasil foram: muitas residências usam apenas antenas internas de baixo custo; existe uma presença agressiva, em quase todas as localidades, de sinais interferentes como ruído impulsivo e distorções por multipercurso; e diferentemente do que ocorre em países de primeiro mundo, as antenas de transmissão das diferentes emissoras não se concentram, necessariamente, em uma única torre, o que impossibilita o uso de uma antena direcional apontada para um único alvo. O relatório final dos testes, contendo estas e outras informações, foi encaminhado para publicação pela Agência Nacional de Telecomunicações (ANATEL).

Em 2003, formaram-se então diversas iniciativas no país com o intuito de propor um modelo para a TV digital brasileira. Com o Laboratório de Sistemas Integráveis da USP protagonizando a exposição da necessidade de uma política estruturante em TV digital no Brasil, esta fase culminou, em seguida, com a publicação do Decreto Presidencial nº

4.901 de 26 de Novembro de 2003 (BRASIL, 2003), que consolidou um consórcio nacional com o nome de Sistema Brasileiro de TV Digital para a definição do sistema a ser adotado no país.

A partir deste decreto teve início a segunda fase, na qual o principal objetivo do consórcio criado era o de promover a integração dos centros de pesquisa brasileiros para apresentar uma solução técnica inovadora de sistema, mantendo e aproveitando a compatibilidade com elementos já padronizados no mercado mundial, em observância com as diretrizes do decreto.

3.1.2 SBTVD e a Decisão do Sistema

Como em 2003 muitos setores já tinham a expectativa de um lançamento próximo para o sistema de TV digital, o SBTVD foi visto com muitas ressalvas e até medo, pois os que se posicionaram contra a iniciativa diziam que este seria uma nova versão do PAL-M, um sistema particular do Brasil que criou dificuldades para a implantação da TV a cores e gerou um legado com o qual se tem que conviver mais de 30 anos depois.

Além disso, alguns requisitos do decreto soaram contraditórios. Falava-se antes de tudo em um sistema como instrumento de inclusão digital e social, principalmente pelo baixo custo, e ao mesmo tempo em tecnologia de ponta e inovações.

No entanto, considerando-se que o sistema digital deveria ter a mesma longevidade e poder de penetração que o sistema analógico, por mais que no lançamento os preços ficassem acima do esperado, logo seriam amortizados, possibilitando a pulverização de transmissores e receptores para o novo sistema. Falava-se também em incentivos fiscais e financiamentos para a produção local de equipamentos, o que também ajudaria na redução de preços.

De qualquer forma, por inovação tecnológica esperava-se a união, com os ajustes necessários, de blocos funcionais consolidados, o que não deveria representar um rompimento com os modelos e o mercado mundiais.

Para organizar as contribuições tecnológicas, a Financiadora de Estudos e Projetos (FINEP) habilitou 83 instituições de ensino, pesquisa e desenvolvimento através da carta convite MC/MCT/FINEP/FUNTTTEL - 01/2004. Em seguida, publicou sete novas cartas organizando as áreas de contribuição, que foram definidas como: Difusão e Acesso; Codificação de Sinal Fonte; Camada de Transporte; Terminal de Acesso; Serviços, Aplicações e Conteúdo; Canal de Interatividade; Middleware.

O LSI teve três propostas aprovadas, um na área de terminal de acesso com o projeto TAR, e outras duas na área de codificação sinal fonte com os projetos dos CODECs H.264/AVC e AAC (Advanced Audio Coding). Além disso, o LSI participou ainda como co-executor de projetos nas áreas de difusão e acesso, canal de interatividade e middleware. Ao todo, foram 21 propostas de redes de parcerias aprovadas e desenvolvidas ao longo de 2005 mediante integração técnica do CPqD e financiamento do Fundo Tecnológico para o Desenvolvimento das Telecomunicações (FUNTTEL). Os resultados desses projetos foram demonstrados publicamente na Escola Politécnica da USP no dia 9 de dezembro de 2005 e consolidados nos relatórios CPqD (2006a) e CPqD (2006b) com propostas para o modelo do sistema brasileiro.

Dentre as publicações do LSI relacionadas aos projetos de 2005 estão a de Carvalho et al. (2007) e a de Hira et al. (2007), que contaram com a participação direta do pesquisador Rogério Pernas Nunes e trataram respectivamente sobre a arquitetura do terminal de acesso de referência para o sistema brasileiro e sobre a interface de software recomendada para o sistema operacional do terminal.

Munido dos subsídios relevantes para a tomada de decisão com relação à escolha do sistema, principalmente pelos resultados de 2005, o governo publicou em 29 de junho de 2006 o Decreto nº 5.820 (BRASIL, 2006), que instituiu então o SBTVD-T, também chamado *Integrated Services for Broadcasting - Terrestrial Brazil* (ISDB-TB). O SBTVD-T tem por base o padrão de transmissão do ISDB-T, incorporando as inovações tecnológicas aprovadas pelo Comitê Interministerial de Desenvolvimento e estabelecendo ainda os formatos HDTV, SDTV e LDTV, a transmissão simultânea para dispositivos fixos, móveis e portáteis e a interatividade.

Finalmente, conforme também estabelecia o decreto, no dia 23 de novembro de 2006 foi criado o Fórum do Sistema Brasileiro de TV Digital Terrestre (Fórum do SBTVD), dando início a terceira e atual fase da TV digital no Brasil.

3.1.3 Fórum SBTVD e Implantação do Sistema

Com a missão de auxiliar e estimular a criação e a melhoria do sistema de transmissão e recepção de sons e imagens digitais no Brasil, propiciando padrão e qualidade compatíveis com as exigências dos usuários, o Fórum SBTVD foi organizado em quatro Módulos de Trabalho, com as seguintes atribuições (página de internet)²:

²Mais informações sobre o Fórum SBTVD podem ser obtidas na página oficial da internet. Disponível em: <<http://www.forumsbtvd.org.br>>. Acesso em: 18 ago. 2008.

Módulo de Mercado: identifica e define os requisitos funcionais, prazos e o custo-alvo dos receptores, coordena a integração, verifica a conformidade das especificações técnicas e analisa e propõe soluções de planejamento para implantação da televisão digital terrestre.

Módulo Técnico: coordena as especificações técnicas (normas) e as atividades de pesquisa e desenvolvimento em todas as áreas pertinentes ao sistema de TV digital. Também coordena a harmonização das especificações técnicas com outras entidades nacionais e internacionais.

Módulo de Propriedade Intelectual: coordena, de um lado, a propriedade intelectual gerada pelo próprio fórum e, de outro, as negociações de *royalties* ligados a tecnologias externas em processo de adoção ou incorporação pelo sistema brasileiro.

Módulo de Promoção: coordena esforços referentes à promoção, divulgação e disseminação do sistema brasileiro de TV digital, sendo responsável por seminários, cursos, edição de jornais, boletins e pelos demais veículos de informação.

Os Módulos de Trabalho são principalmente compostos por membros das emissoras do serviço de radiodifusão de sons e imagens, por fabricantes de equipamentos de transmissão de sinais de televisão terrestre, por fabricantes de equipamentos de recepção de sinais da televisão terrestre, pelas indústrias de software e pelas entidades de ensino e pesquisa.

O Módulo Técnico, subdividido em sete Grupos de Trabalho distintos, um para cada grande bloco sistêmico, elaborou as normas essenciais para que o SBTVD-T entrasse finalmente em operação no dia 2 de dezembro de 2008 com inauguração realizada na cidade de São Paulo.

O pesquisador Rogério Pernas Nunes tem participado, desde a inauguração do Fórum SBTVD, do grupo de Codificação Sinais Fonte. Contribuiu ativamente durante esse tempo na elaboração da norma de Codificação de Vídeo e de seu Guia de Operação.

O cronograma de implantação da TV digital terrestre no Brasil está seguindo prazos estipulados na Portaria do Ministério das Comunicações nº 652 de 10 de outubro de 2006. Na realidade, a portaria especifica os prazos para solicitação de consignação dos canais digitais e as etapas seguintes, de onde é possível derivar os intervalos para início das transmissões de canais consignados por região, conforme a Figura 3.2, que mostra desde a data de inauguração do sistema na cidade de São Paulo.

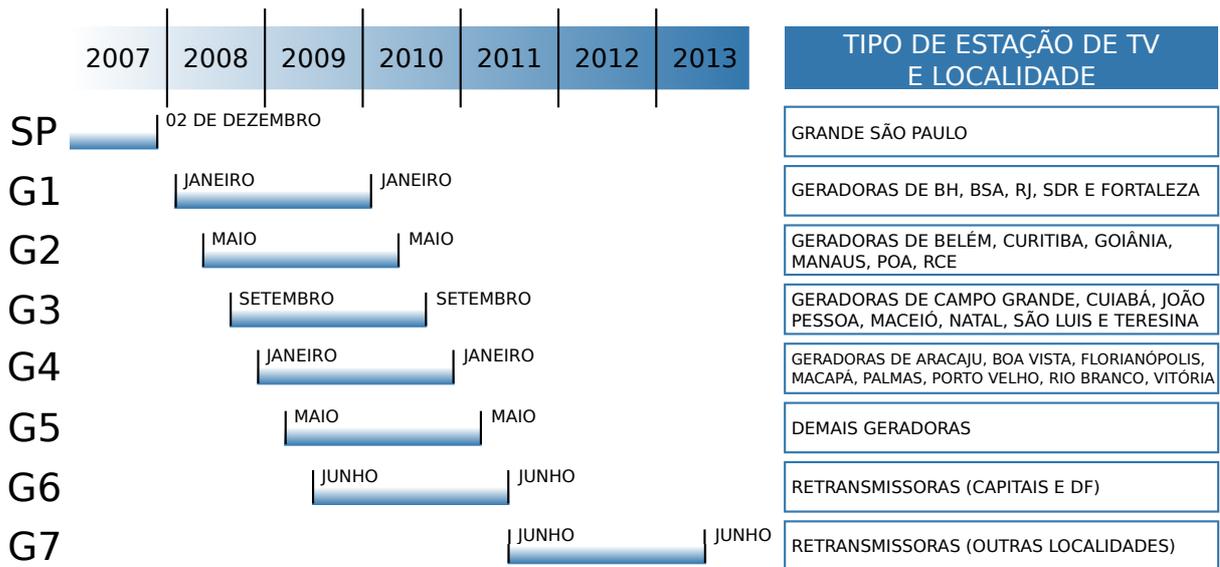


Figura 3.2: Cronograma de implantação da TV digital terrestre no Brasil

3.2 Visão Geral das Normas do SBTVD-T

A Figura 3.3 mostra o diagrama de blocos do SBTVD-T. Cada um dos sete blocos sistêmicos está associado a um Grupo de Trabalho respectivo no Módulo Técnico do Fórum SBTVD. Os Grupos de Trabalho fazem parte também de uma comissão especial da Associação Brasileira de Normas Técnicas (ABNT) responsável pela publicação oficial das normas.

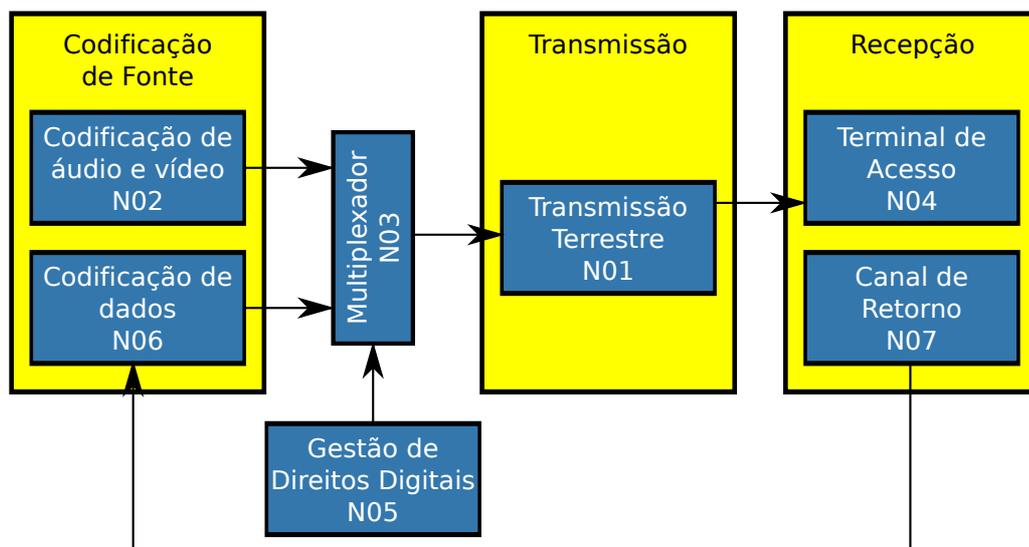


Figura 3.3: Normas que compõem o SBTVD-T

O identificador (N0X) dentro de cada bloco é a referência interna dos Grupos de Trabalho para as normas elaboradas, sendo que algumas delas são formadas por mais de

uma parte. A Tabela 3.1 apresenta a lista de normas (página de internet)³ já aprovadas e publicadas oficialmente ou que ainda estão em elaboração pelos Grupos de Trabalho. A tabela apresenta também os principais documentos que foram utilizados como referência na elaboração de cada norma e o nome oficial de referência de cada norma conforme cadastro na ABNT:

Tabela 3.1: Lista de normas do SBTVD-T

Norma SBTVD-T	Referências	Nome na ABNT
N01 Sistema de transmissão	ARIB STD-B31	ABNT NBR 15601
N02 Codificação de vídeo, áudio e multiplexação		
Parte 1 Codificação de vídeo	ITU-T H.264 ISO/IEC 14496-10	ABNT NBR 15602-1
Parte 2 Codificação de áudio	ISO/IEC 14496-3	ABNT NBR 15602-2
Parte 3 Multiplexação de sinais	ITU-T H.222 ISO/IEC 13818-1	ABNT NBR 15602-3
N03 Multiplexação e serviços de informação (SI)		
Parte 1 SI do sistema de radiodifusão	ARIB STD-B10	ABNT NBR 15603-1
Parte 2 Estrutura de dados e definições da informação básica de SI	ARIB STD-B10	ABNT NBR 15603-2
Parte 3 Sintaxes e definições de informação estendida do SI	ARIB STD-B10	ABNT NBR 15603-3
N04 Receptores	ARIB STD-B10	ABNT NBR 15604
N05 Tópicos de segurança		
Parte 1 Controle de cópias	ARIB TR-B14	ABNT NBR 15605-1
Parte 2 Mecanismos de segurança para aplicativos		EM ELABORAÇÃO
N06 Codificação de dados e especificações de transmissão para radiodifusão digital		
Parte 1 Codificação de dados	ARIB STD-B24 ETSI TS 101812 ETSI TS 102819	ABNT NBR 15606-1
Parte 2 Ginga NCL para receptores fixos e móveis	ARIB STD-B24 ETSI TS 101812 ETSI TS 102819	ABNT NBR 15606-2
Parte 3 Especificação de transmissão de dados	ARIB STD-B24 ETSI TS 101812 ETSI TS 102819	ABNT NBR 15606-3
Parte 4 Ginga-J		EM ELABORAÇÃO
Parte 5 Ginga NCL para receptores portáteis	ARIB STD-B24 ETSI TS 101812 ETSI TS 102819	ABNT NBR 15606-5
N07 Canal de interatividade		
Parte 1 Protocolos, interfaces físicas e de software	ARIB STD-B21 ARIB STD-B24 ETSI TS 101	ABNT NBR 15607-1
Parte 2 Dispositivos externos		EM ELABORAÇÃO
Parte 3 Interface de configuração para as tecnologias de acesso		EM ELABORAÇÃO

Além das normas listadas, o Fórum SBTVD também é responsável pela elaboração de Guias de Operação que auxiliam na implementação do sistema. Estes guias estão

³ABNT. Na página de internet da ABNT estão disponíveis todas as normas do SBTVD para consulta. Disponível em: <http://www.abnt.org.br/tvdigital>. Acesso em: 27 dez. 2008.

disponíveis sob o número ABNT NBR 15608-X, mas não devem ter efeito normativo, servindo como referência de melhores práticas e parâmetros para implementar as normas associadas e operar o sistema digital. Existem ainda outras duas normas especiais não listadas, que também encontram-se em fase de elaboração: ABNT NBR 15609 - Suíte de testes e ABNT NBR 15610 - Ensaio para receptores. Estas duas normas deverão servir para homologação do sistema e dos receptores.

Dentre todas as normas citadas, as particularmente relevantes para este trabalho, que serão apresentadas a seguir, são: N01, N02 Parte 1 e N02 parte 3.

3.2.1 Sistema de Transmissão (N01)

Seguindo as normas japonesas como referência, no SBTVD-T foi especificada uma técnica com modulação *Orthogonal Frequency Division Multiplexing* (OFDM) para o sistema de transmissão (ABNT, 2007a). Esta técnica, denominada BST-OFDM, foi desenvolvida para permitir a transmissão da TV digital terrestre com modulação hierárquica e com maior flexibilidade de serviços (JUNIOR, 2008). O canal de 6 MHz é dividido em 14 segmentos de 428,57 KHz, dos quais 13 segmentos podem ser utilizados na transmissão de até três serviços distintos e simultâneos, com diferentes características de modulação e robustez. O segmento restante é dividido entre duas metades que funcionam como bandas de guarda no início e no final do canal.

A Figura 3.4 mostra um canal de 6 MHz como exemplo, onde se observam as bandas de guarda nas laterais e a ocupação dos segmentos por três serviços distintos: um serviço HDTV ocupando 9 segmentos com modulação 64-QAM para recepção fixa; um serviço SDTV ocupando 3 segmentos com modulação 16-QAM para recepção fixa e móvel; e um serviço LDTV de 1 segmento (*one-seg*) com modulação QPSK para recepção portátil.

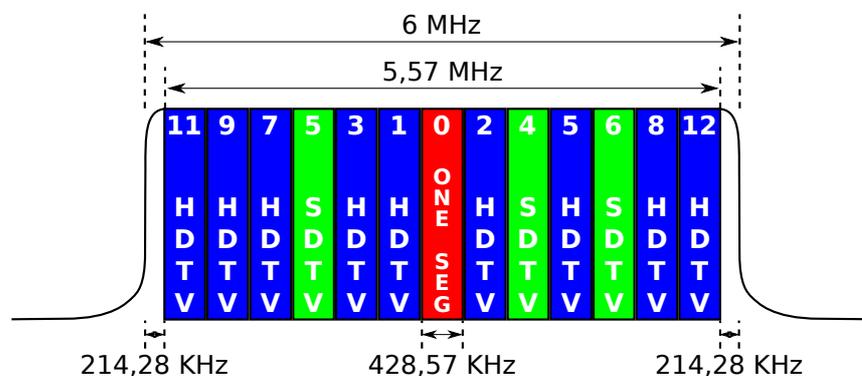


Figura 3.4: Exemplo de uso do canal de 6 MHz no SBTVD-T e bandas de guarda

Na modulação BST-OFDM, o segmento central realmente concentra todas as portadoras do serviço *one-seg* como mostra a figura. As portadoras dos outros serviços, no entanto, estão dispersas e intercaladas pelo espectro.

No exemplo anterior, em função dos demais parâmetros configuráveis no sistema de modulação por camada hierárquica (intervalo de guarda e código convolucional), as taxas de bits por camada podem variar conforme a Tabela 3.2.

Tabela 3.2: Taxas máxima e mínima por serviço do canal de exemplo

Serviço	Modulação	Taxa mínima (kbps)	Taxa máxima (kbps)
LDTV (1 segmento)	QPSK	280,85	595,76
SDTV (3 segmentos)	16-QAM	1685,13	3574,56
HDTV (9 segmentos)	64-QAM	7583,13	16085,52

O sistema de transmissão permite ainda que se escolha um entre os três modos disponíveis, sendo que cada modo representa um espaçamento distinto entre portadoras BST-OFDM. No modo 1, os espaços entre portadoras são de aproximadamente 4 kHz, no modo 2, aproximadamente 2 kHz, e no modo 3, aproximadamente 1 kHz. O modo mais adequado para operação deve ser escolhido tanto em função da robustez ao efeito *Doppler* para recepção móvel e portátil quanto em função da distância entre transmissores em uma rede com frequência única. Essa escolha não afeta a taxa de transmissão, desde que sejam mantidos os outros parâmetros do sistema.

No SBTVD-T, outro recurso importante herdado do ISDB-T foi a decalagem (deslocamento do centro das portadoras BST-OFDM em 1/7 MHz para mais), que proporciona menor interferências dos canais digitais em canais analógicos adjacentes.

A Figura 3.5 apresenta o diagrama de blocos da transmissão no SBTVD-T. Os sinais fonte de áudio e vídeo (à esquerda) são codificados de acordo com o tipo de serviço (HDTV, SDTV, LDTV), formando um fluxo de bits, denominado Fluxo de Transporte (*Transport Stream*) (TS), de áudio e vídeo para cada serviço. Estes fluxos são multiplexados (MUX), juntamente com os fluxos de bits de dados e as tabelas de informação, que serão apresentadas na Sessão 3.2.3, em um único fluxo que é encaminhado ao re-multiplexador (RE-MUX) (UEHARA, 2006). O re-multiplexador insere tabelas adicionais de transmissão, preparando os dados de entrada dos moduladores. Cada modulador corresponde a um dos serviços ou camada do sistema e possui parâmetros independentes como visto anteriormente. As etapas finais são comuns a todas as camadas que estão sendo transmitidas e que já foram devidamente configuradas e tratadas. Este tipo de transmissão é chamado transmissão hierárquica.

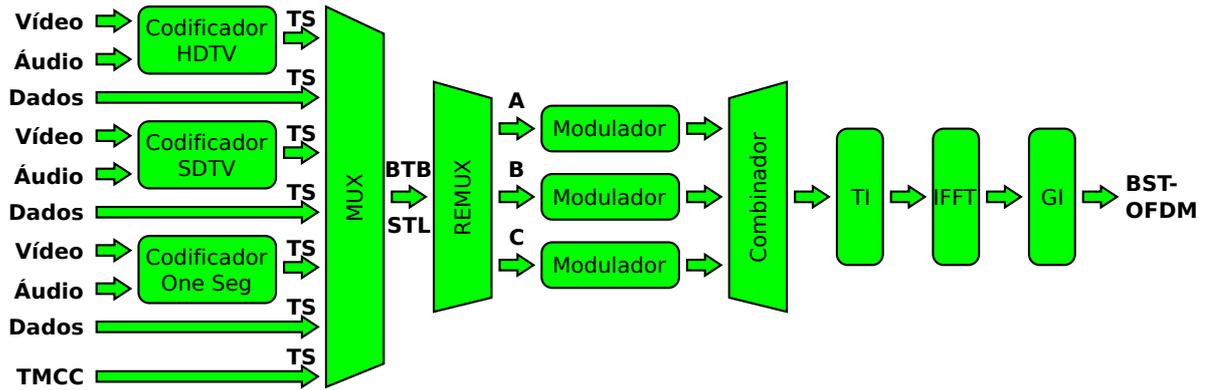


Figura 3.5: Diagrama de blocos de transmissão do SBTVD-T

A Figura 3.6 mostra o espectro alocado pela ANATEL para a radiodifusão da televisão digital no Brasil. Os números na parte superior da Figura indicam os intervalos dos canais físicos VHF e UHF que pertencem ao sistema e os números na parte inferior indicam os intervalos de frequência em MHz que compreendem os canais físicos. O canal 37 (frequência 611 MHz) não é utilizado para a TV Digital.

7	13	14	37 Canais	69
VHF		UHF		
174	216	470	Frequências (MHz)	806

Figura 3.6: Espectro alocado pela ANATEL para o SBTVD

3.2.2 Codificação de Vídeo (N02 Parte 1)

Durante os estudos de 2005 foram avaliados os dois padrões de codificação mais adequados para a radiodifusão: o MPEG-2, que já era largamente utilizado no mundo, e o H.264/AVC, que apesar de recente alguns países já ensaiavam adotar, principalmente em sistemas fechados de televisão, menos presos a legado. Somente o MPEG-2 foi avaliado também quanto ao suporte a escalabilidade, pois para o H.264/AVC as ferramentas relacionadas ainda estavam em estudo e não haviam sido acrescentadas ao padrão.

Na época verificou-se que a escalabilidade espacial direta da resolução de um conteúdo em SDTV para HDTV com MPEG-2 não apresentava ganho quando comparada ao envio de conteúdo SDTV e HDTV simultaneamente no mesmo fluxo. O uso da escalabilidade somente se justificava caso houvesse necessidade de resoluções intermediárias, o que não era o caso.

Por outro lado, a codificação H.264/AVC, mesmo sem escalabilidade, já se mostrava

vantajosa. Realmente cumpria a expectativa de reduzir em média pela metade a taxa de bits em comparação com o MPEG-2, tendo como única ressalva à época a baixa disponibilidade de chips no mercado e os preços, principalmente para conteúdo HDTV.

Quando em 2006 os trabalhos de especificação do sistema pelo Fórum SBTVD tiveram início, o cenário havia mudado. A oferta de chips de decodificação de H.264/AVC em alta resolução estava aumentando no mercado e os preços já começavam a apresentar sinais de queda. Por esses motivos, diferentemente da codificação de vídeo MPEG-2 utilizada nas normas japonesas, como resultado das inovações propostas no SBTVD-T foi adotado o H.264/AVC, ainda sem suporte à escalabilidade.

A versão final da norma de codificação de vídeo (ABNT, 2007b) especifica algumas restrições e obrigações relativas ao uso de parâmetros e elementos do H.264/AVC. Alguns elementos de sintaxe que carregam informações adicionais, Informações Suplementares de Enriquecimento (*Supplemental Enhancement Information*) (SEI) e Informações de Usabilidade de Vídeo (*Video Usability Information*) (VUI), devem estar presentes em um fluxo de bits compatível com o sistema e, além disso, os receptores devem interpretar estes elementos e realizar o processamento necessário.

Os elementos VUI obrigatórios no sistema são: *aspect_ratio_idc* (índice de relação de aspecto do pixel utilizado); *colour_primaries*, *transfer_characteristics*, e *matrix_coefficients* (informação de colorimetria); *time_scale*, *num_units_in_ticks* e *fixed_frame_rate_flag* (informação de taxa de quadros); *pic_struct_present_flag* (indica a presença de informação mais detalhada sobre a temporização de cada figura da seqüência em mensagem SEI).

E o único elemento SEI obrigatório é o *picture timing SEI message* (*pic_timing*), que informa detalhes sobre a temporização de cada figura da seqüência.

Existem ainda dois elementos SEI opcionais no sistema: *pan and scan rectangle SEI message*, que define dinamicamente os retângulos de interesse na seqüência decodificada, para adaptar a saída de vídeo do receptor a monitores com características diferentes das do vídeo original; e *user data registered by ITU-T Recommendation T.35 SEI message*, que possui objetivo similar ao do *pan and scan*, porém é menos flexível e, portanto, também menos custoso em termos de processamento.

O conceito de seqüência de vídeo é definido na norma brasileira como um conjunto de unidades de acesso, sendo que a primeira unidade deve ser do tipo *Instantaneous Decoding Refresh* (IDR) e a última deve terminar com uma unidade NAL do tipo *end_of_seq*. Além disso, um fluxo de vídeo é definido como um conjunto de seqüências conforme a definição anterior, terminado por uma unidade do tipo *end_of_stream*.

A norma brasileira define ainda o conceito de Ponto de Acesso Aleatório (*Random Access Point*) (RAP), que é uma unidade de acesso a partir da qual o vídeo pode ser decodificado sem dependência das informações anteriores. O máximo intervalo permitido entre unidades do tipo RAP é de 5 s, embora o intervalo recomendado para operação seja inferior a 1 s para que o início da decodificação não fique muito demorado na troca de canal, por exemplo.

Ao final da norma, são apresentadas diversas recomendações para o chaveamento contínuo de conteúdo, que definem os procedimentos na transmissão e na recepção para permitir a exibição contínua de imagens pelo receptor, quando ocorre o chaveamento entre formatos de vídeo diferentes ou quando ocorrem modificações nos parâmetros que definem a seqüência de vídeo.

Quanto aos principais parâmetros que definem as características da seqüência de vídeo, a norma brasileira encontra-se dividida em duas partes: serviços *full-seg* (relativa a vídeos SDTV e HDTV), com foco em recepção fixa; e serviços *one-seg* (relativa aos vídeos LDTV), com foco em recepção portátil.

3.2.2.1 Parâmetros Específicos dos Serviços *Full-seg*

Fluxo de bits codificado de vídeo para serviço *full-seg* deve pertencer ao perfil Main ou, preferencialmente, ao perfil High e deve ser enquadrado em nível igual ou inferior ao 4.0.

As restrições aplicáveis aos parâmetros de resolução de luminância e taxa de quadros são exibidas respectivamente na Tabela 3.3 e na Tabela 3.4.

Tabela 3.3: Resoluções de luminância com suporte obrigatório para SDTV e HDTV

Tipo de resolução	Resolução de luminância (horizontal x vertical)	Razão de aspecto	<i>aspect_ratio_idc</i>
SD	720 x 480	4:3	3
		16:9	5
HD	1280 x 720	16:9	1
HD	1920 x 1080	16:9	1

Tabela 3.4: Taxas de quadros com suporte obrigatório para SDTV e HDTV

Taxa de quadros	Tipo de seqüência	<i>time_scale</i>	<i>num_units_in_tick</i>
30000/1001	Progressiva	60000	1001
30000/1001	Entrelaçada	60000	1001
60000/1001	Progressiva	120000	1001

Apesar de permitir sinalização de qualquer tipo de colorimetria no fluxo de bits, a

norma recomenda a utilização da ITU-R Recommendation BT.601-6 para fluxos SDTV e da ITU-R Recommendation BT.709-5 para fluxos HDTV, que se refletem nos seguintes valores de parâmetros para os fluxos, conforme Tabela 3.5:

Tabela 3.5: Parâmetros de colorimetria recomendados para SDTV e HDTV

Parâmetro	SDTV	HDTV
<i>colour primaries</i>	6	1
<i>transfer characteristics</i>	6	1
<i>matrix coefficients</i>	6	1

3.2.2.2 Parâmetros Específicos dos Serviços *One-seg*

Fluxo de bits codificado de vídeo para serviço *one-seg* deve pertencer ao perfil Baseline e deve ser enquadrado em nível igual ou inferior ao 1.3.

As restrições aplicáveis aos parâmetros de resolução de luminância e taxa de quadros são exibidas, respectivamente, na Tabela 3.6 e na Tabela 3.7.

Tabela 3.6: Resoluções de luminância com suporte obrigatório para LDTV

Formato	Resolução de luminância (horizontal x vertical)	Razão de aspecto	<i>aspect_ratio_idc</i>
SQVGA	160 x 120	4:3	1
SQVGA	160 x 90	16:9	1
QVGA	320 x 240	4:3	1
QVGA	320 x 180	16:9	1
CIF	352 x 288	4:3	2

Tabela 3.7: Taxas de quadros com suporte obrigatório para LDTV

Taxa de quadros	Tipo de seqüência	<i>time_scale</i>	<i>num_units_in_tick</i>
5 Hz	Progressiva	5000	1001
10 Hz	Progressiva	10000	1001
12 Hz	Progressiva	120000	1001
15 Hz	Progressiva	150000	1001
24 Hz	Progressiva	240000	1001
30 Hz	Progressiva	300000	1001

Além disso, a norma de codificação de vídeo do SBTVD-T não permite o uso das ferramentas FMO, ASO e RS definidas no perfil Baseline da norma H.264/AVC.

3.2.3 Sistemas de Multiplexação de Sinais (N02 Parte 3)

A norma N02 Parte 3 (ABNT, 2007d) é na verdade um resumo das três partes em que está dividida a N03 (ABNT, 2007e; ABNT, 2007f; ABNT, 2007g). Ela aborda o mecanismo essencial de encapsulamento e de combinação dos fluxos de áudio, vídeo, dados e tabelas associadas em um único fluxo de transporte. Com esta norma é possível criar um fluxo de bits simples válido para o SBTVD, com o mínimo de informação que permite navegação entre canais e serviços, bem como entre múltiplos fluxos de áudio e vídeo de um mesmo serviço.

A principal referência para esta norma é a norma de camada de transporte, criada em conjunto pela ITU e pela ISO/IEC, respectivamente com os nomes de ITU-T H.222 e ISO/IEC 13818-1 nessas organizações. A camada de transporte especificada nessas normas é também chamada camada de sistema ou MPEG-2 Sistemas e suporta seis funções básicas:

1. Sincronização de múltiplos fluxos comprimidos na decodificação.
2. A combinação de múltiplos fluxos comprimidos em um único fluxo de bits.
3. A inicialização de memória buffer para começar uma decodificação.
4. Gerenciamento contínuo da memória buffer.
5. Identificação temporal.
6. Combinação e sinalização de vários componentes num fluxo de sistema.

Um fluxo de bits em conformidade com o MPEG-2 Sistemas pode ser um Fluxo de Transporte (*Transport Stream*) (TS) ou um Fluxo de Programa (*Program Stream*) (PS). O primeiro é mais adequado para a transmissão em redes susceptíveis a erros e o segundo é mais adequado para armazenamento local. Além disso, o TS suporta sinalização e conteúdos de múltiplos programas num mesmo fluxo, enquanto o PS é restrito a um único programa por fluxo. O sistema de TV digital, por essas características, exige que o TS seja utilizado, não permitindo o uso de PS na camada de sistema.

A Figura 3.7 mostra as etapas de construção de um TS de TV digital a partir dos sinais de áudio, vídeo e dados.

Primeiro, cada fluxo comprimido gerado por um codificador, ainda sem camada de transporte e chamado Fluxo Elementar (*Elementary Stream*) (ES), é segmentado em quadros, que são encapsulados em pacotes do tipo *Packetized Elementary Stream* (PES).

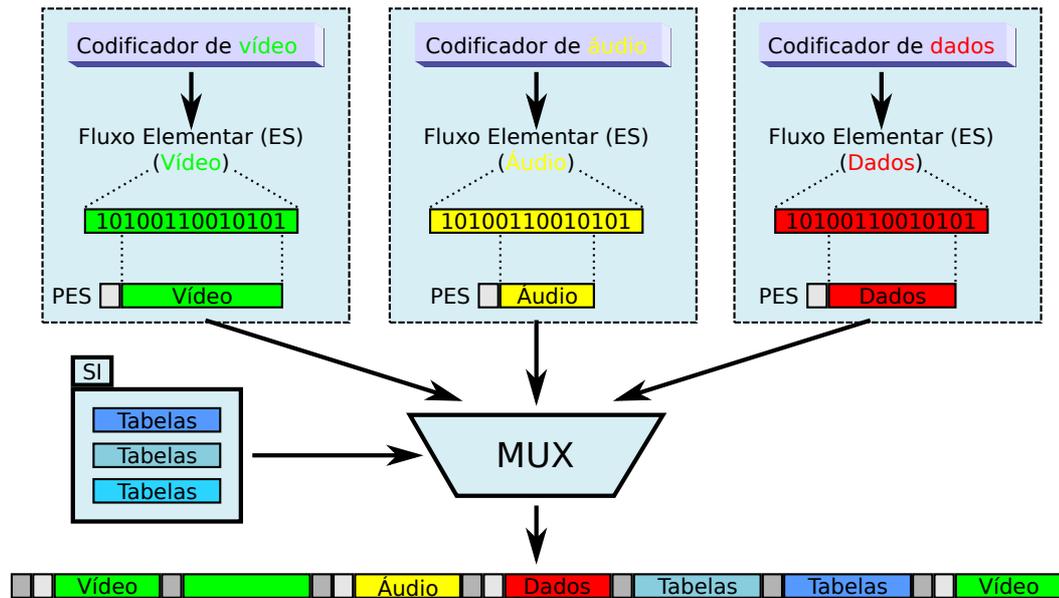


Figura 3.7: Relação entre sinais fonte, pacotes PES, tabelas SI e fluxo TS

No caso do vídeo, o ES é um fluxo serial de unidades do tipo NAL.

Um pacote PES tem tamanho variado, normalmente da ordem de kbytes, e é formado por um cabeçalho e um quadro de informação de áudio, vídeo ou dados, como mostra a Figura 3.8 em detalhe.

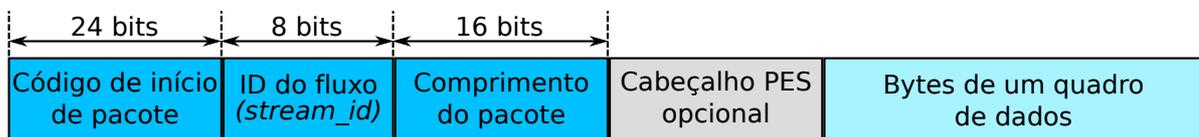


Figura 3.8: Estrutura de um pacote PES

Os valores possíveis para o campo *stream_id* são especificados pela norma MPEG-2 Sistemas. Para pacotes de vídeo, os valores possíveis devem ser da forma ‘1110xxxx’ em binário, onde ‘x’ pode assumir valor ‘0’ ou ‘1’. Neste caso os quatro primeiros bits identificam o tipo de fluxo como vídeo e os quatro últimos servem para numeração. Além disso, as informações de temporização do quadro transportado estão no “Cabeçalho PES opcional”.

Em seguida, os PESs são segmentados em pacotes TS e combinados entre si e com as tabelas auxiliares, formando um fluxo único, representado à saída do MUX na Figura 3.7.

O TS é formado por pacotes com tamanho fixo de 188 bytes, divididos entre um cabeçalho próprio, que pode ter tamanho variado, como mostra a Figura 3.9, e segmentos de pacotes PES ou tabelas que ocupam o restante dos bytes do pacote.

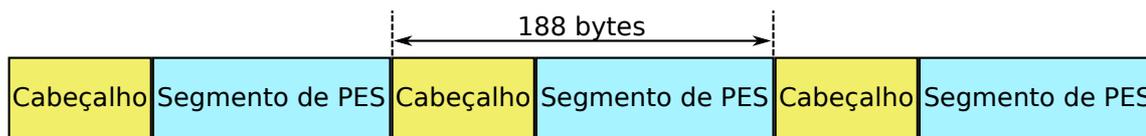


Figura 3.9: Estrutura de pacotes de um TS

Algumas tabelas transportadas no TS são definidas no MPEG-2 Sistemas, onde são denominadas Informação Específica de Programa (*Program Specific Information*) (PSI). O restante das tabelas é definido pelo próprio sistema de TV digital, que referencia a todas genericamente como Informação de Sistema (*System Information*) (SI).

O cabeçalho do pacote TS possui um campo para identificação com 13 bits, denominado PID. Alguns PIDs são pré-definidos no sistema e outros são derivados a partir das tabelas SI.

A Tabela 3.8 lista as principais tabelas SI do sistema e o PID do pacote TS associando, caso este seja fixo, ou o intervalo de valores permitidos caso contrário.

Tabela 3.8: Tabelas principais do TS no SBTVD-T

Tabela	PID	Descrição
PAT	0x0000	<i>Program Association Table</i> : Associa número de programa com o PID onde se encontra a tabela PMT que o descreve.
PMT	0x0011 - 0x1FFE	<i>Program Map Table</i> : Associa um número de programa a seus elementos, ou seja, aos PIDs dos fluxos de áudio, vídeo e dados que o formam.
NIT	0x0010	<i>Network Information Table</i> : Carrega parâmetros físicos da rede de serviços. Exemplo: frequência de sintonia do canal.

Para finalizar o resumo dos principais pontos da N02 Parte 3, segue lista com os principais pontos de diferença, relativos à multiplexação, entre a norma brasileira e a japonesa:

Obrigatoriedade no envio do Canal virtual: Canal virtual é uma numeração única para cada emissora, idêntica à numeração analógica para aquelas que já operavam neste sistema. O canal virtual também é chamado de *remote_control_key_id*, que é um campo do descritor⁴ *TS_information_descriptor* localizado na tabela NIT.

Classificação Indicativa: Definida conforme portaria nº 1100 do Ministério da Justiça Brasileiro.

⁴Descritores são estruturas de dados que podem ser incluídas nas tabelas SI.

Classificação de gênero e subgênero dos programas: Foram adicionados outros itens, em ralação à norma japonesa, para adequação à cultura brasileira.

Tabelas de referência de tempo: O Japão não apresenta diferentes fusos-horários e utiliza o JST (*Japanese Standard Time*), definido a partir do horário de Tóquio, ao invés do UTC (*Universal Time Coordinated*), como no Brasil para representar o horário em seu receptor.

Código de área: Identifica as áreas cobertas pelo sinal de TV digital. Devido à grande diferença territorial entre o Brasil e o Japão, o código especificado no Brasil baseia-se no levantamento efetuado pelo IBGE para localização dos estados e microrregiões.

Identificação da rede de origem: O Brasil possui, atualmente, 496 estações geradoras de sinal de TV. O *original_network_id* é um número atribuído pela ANATEL, que deve ser transmitido para identificação de cada estação.

Identificação do serviço: O campo *service_id* também foi modificado no sistema brasileiro. Ele é responsável por identificar unicamente a estação geradora (emissora), o tipo e o número do serviço transmitido. Cada canal pode transmitir até 8 serviços simultâneos.

3.3 Conclusão

Este capítulo apresentou o Sistema Brasileiro de TV Digital, inicialmente relatando seu histórico sob o ponto de vista do autor, que o divide em três fases, tendo participado a partir da segunda fase no processo de especificação e implantação.

Foram apresentados, em seqüência, a estrutura organizacional e um quadro resumo das normas que compõem o SBTVD até o início de 2009.

Sob o ponto de vista de transmissão, verificou-se que o SBTVD apresenta um sistema de modulação flexível e avançado, que permite a exploração de um canal digital através de diversas formas de configuração.

A parte de especificação da codificação do sistema apresentou a ferramenta H.264/AVC, escolhida de forma pioneira, e suas principais características particulares para o SBTVD. Cabe ressaltar a Sessão 3.2.2.1, que apresentou os parâmetros específicos para os serviços orientados a receptores fixos.

Por fim, verificou-se que a camada de transporte, baseada na parte sistêmica do

MPEG-2, permite a multiplexação de áudio, vídeo e dados e, através de tabelas de informação, permite a organização do conteúdo que carrega.

Pelas características identificadas no SBTVD ao longo do capítulo, verifica-se que este sistema pode sofrer evoluções futuras. Especificamente no próximo capítulo será discutido como o SBTVD poderá evoluir na questão da escalabilidade dos formatos de vídeo.

4 Considerações sobre a Escalabilidade de Vídeo para o Sistema Brasileiro de TV Digital

Este capítulo tem por objetivo apresentar considerações sobre a escalabilidade de vídeo para o SBTVD. Para tanto, serão introduzidos alguns trabalhos atuais relacionados tanto à escalabilidade quanto à radiodifusão, que apresentam resultados importantes a serem considerados no direcionamento da discussão seguinte, específica do SBTVD. Algumas das considerações abordadas serão utilizadas na especificação de requisitos para montagem de um ambiente de investigação experimental.

Espera-se com a implementação do ambiente experimental, dentre outros resultados, evidenciar algumas das premissas da escalabilidade, bem como investigar o estado-da-arte de ferramentas tanto de codificação de vídeo quanto de transmissão e análise de sinais digitais de televisão.

Como premissa deste trabalho, foi considerado como principal requisito para os cenários de evolução do SBTVD a preservação dos receptores legados. O método mais simples que permitiria a evolução nos formatos de vídeo disponíveis seria o *simulcast*, que será utilizado como base de comparação.

Apesar deste trabalho não ter o objetivo de confrontar a escalabilidade do H.264/AVC com outras soluções de escalabilidade, o *simulcast* será particularmente abordado por ser o método mais simples que permite a evolução de um sistema sem comprometer o legado e consiste na transmissão simultânea do mesmo conteúdo tanto em formato antigo quanto novo. Este nome se aplica, por exemplo, ao período de migração do sistema analógico de radiodifusão para o digital, durante o qual os dois sistemas permanecem em operação, transmitindo o mesmo conteúdo. Espera-se que a escalabilidade apresente vantagem significativa se comparada ao *simulcast*.

Os resultados de qualidade deste trabalho serão apresentados com métricas objetivas

para análises comparativas. No entanto, conforme apresentado no Capítulo 2, quando se deseja uma medida absoluta para análise de qualidade de vídeo, é necessário realizar avaliações subjetivas. Dessa forma, entre os trabalhos futuros, serão estudados métodos como os utilizados em EBU (2008) para extensão da montagem experimental, a fim de permitir análises subjetivas de vídeo.

4.1 **Trabalhos Correlatos**

A norma de codificação de vídeo H.264/AVC e sua extensão de escalabilidade foram resultados de investigações normalmente publicadas em artigos científicos. No entanto, encontra-se literatura importante sobre a aplicação destas tecnologias em sistemas de radiodifusão em relatórios técnicos de entidades como DVB, EBU, NHK, ISO/IEC e ITU.

Nesta seção são apresentados alguns aspectos e resultados do Relatório de Teste de Verificação do SVC, do projeto Super Hi-Vision e das recentes iniciativas de análise do SVC em cenários de aplicações diversas, resumida pela EBU, e de extensão das normas europeias de transmissão digital para suporte à escalabilidade de vídeo.

4.1.1 **Resultados de Testes de Escalabilidade**

O grupo JVT apresentou em Janeiro de 2007 resultados de testes de verificação (JVT, 2007) que foram conduzidos para medir o desempenho do SVC. Os testes foram realizados em condições apropriadas para uma faixa de cenários de aplicações baseadas em vídeo progressivo. Entre os cenários, foi contemplada a transmissão em HDTV, através de escalabilidade espacial de 720p@50qps¹ para 1080p@50qps.

Para as avaliações de desempenho, foram realizados testes subjetivos, comparando fluxos de vídeo SVC com fluxos simples codificados em H.264/AVC sem escalabilidade. Os testes subjetivos foram conduzidos seguindo recomendações internacionais relevantes (notadamente a ITU-R BT.500-11) em um ambiente controlado e com grande número de voluntários.

Os resultados, de forma geral, evidenciam que as taxas de bits dos fluxos codificados pelo SVC, com camada base de 720p e camada de enriquecimento de 1080p, apresentam

¹Nesta forma mais completa de apresentar o formato de vídeo, além da resolução e do tipo de seqüência (entrelaçada ou progressiva), fica explícita também a taxa de quadros. Neste caso, 50 quadros por segundo. No caso de vídeo entrelaçado, a taxa de campos é o dobro da taxa de quadros. Pode-se suprimir a unidade (qps), deixando-a implícita.

um aumento típico igual ou menor a 10% em comparação às taxas de bits dos fluxos codificados em camada simples de 1080p com o H.264/AVC. Além disso, os resultados evidenciam que as taxas de bits dos fluxos SVC apresentam economia de até 40% em relação à soma das taxas dos fluxos codificados individualmente em 720p e 1080p com o H.264/AVC, na situação de *simulcast*.

A Tabela 4.1 apresenta a consolidação para o cenário de radiodifusão das seqüências avaliadas e os resultados obtidos em taxa de bits, PSNR e Índice de Opinião Média (*Mean Opinion Score*) (MOS)², sendo este último um índice com valor entre 0 e 10, refletindo a qualidade do vídeo na avaliação subjetiva dos voluntários. As seqüências originais empregadas nos testes não estão disponíveis publicamente para que os testes sejam reproduzidos. Estão disponíveis apenas os arquivos de configuração para o codificador de referência, os fluxos já codificados, e arquivos de execução em lote (*shell scripts*) para automatizar algumas tarefas de codificação e decodificação.

Tabela 4.1: Resultados dos testes de escalabilidade do JVT para HDTV

Seqüência Codificada	Taxa de bits (kbps)	PSNR (dB)	Qualidade (MOS)
AlohaWave_AVC_720p50@4Mbps	3987,62	36,55	7,55
AlohaWave_AVC_1080p50@6Mbps	6016,89	32,18	6,57
AlohaWave_SVC_720p50@4Mbps	3988,93	36,55	8,00
AlohaWave_SVC_720p50@4Mbps-1080p50@6Mbps	5946,83	32,13	7,63
CrowdRun_AVC_720p50@8Mbps	7922,65	30,59	6,90
CrowdRun_AVC_1080p50@13Mbps	13122,32	29,71	6,52
CrowdRun_SVC_720p50@8Mbps	7961,59	30,58	7,19
CrowdRun_SVC_720p50@8Mbps-1080p50@13Mbps	13070,37	28,84	6,38
Seeking_AVC_720p50@4Mbps	3990,78	31,24	7,67
Seeking_AVC_1080p50@6Mbps	5981,46	29,81	6,52
Seeking_SVC_720p50@4Mbps	3977,99	31,18	6,67
Seeking_SVC_720p50@4Mbps-1080p50@6Mbps	6020,58	29,38	6,05
Umbrella_AVC_720p50@6Mbps	6038,14	31,75	8,19
Umbrella_AVC_1080p50@10Mbps	10071,68	30,25	7,71
Umbrella_SVC_720p50@6Mbps	6009,41	31,74	7,80
Umbrella_SVC_720p50@6Mbps-1080p50@10Mbps	10089,68	29,66	7,43

Nesta avaliação, foram consideradas taxas de bits próximas entre as camadas equivalentes dos fluxos simples e dos fluxos com escalabilidade. Pode-se observar na Tabela 4.1 que para a camada base em 720p, houve variação desprezível na qualidade objetiva, com variação pequena ora para mais, ora para menos, no valor do índice de qualidade subjetiva. Já a comparação entre o fluxo simples em 1080p e o fluxo enriquecido, codificado

²As faixas do índice de qualidade subjetiva estão divididas em: [0-2] - Ruim; [2-4] - Pobre; [4-6] - Satisfatório; [6-8] - Bom; [8-10] - Excelente.

com SVC, mostra variação ainda pequena, porém mais perceptível da qualidade objetiva e também da subjetiva, com exceção para a primeira seqüência. Alguns artefatos de codificação tornaram-se visíveis nos fluxos de alta resolução, refletindo em menores valores de qualidade objetiva e subjetiva em relação aos fluxos equivalentes de menores resoluções.

4.1.2 **Estágio Recente do Sistema Super Hi-Vision**

Ainda que a expectativa inicial de lançamento do sistema Super Hi-Vision estivesse em torno de 2025, suas transmissões experimentais devem ser antecipadas para que comecem em 2015 pelo cronograma atual do projeto. Como esse sistema pode ser o grande direcionador da tecnologia de produção e transmissão de vídeo nos próximos anos, embora ele não tenha, até o momento, contemplado codificação escalável de vídeo, é apresentado seu estágio de desenvolvimento atual e suas expectativas como referência (SUGAWARA, 2008; SHISHIKUI; FUJITA; KUBOTA, 2009).

Em artigo publicado no segundo semestre de 2007 (SAKAIDA et al., 2007), os autores (cientistas e técnicos da NHK) apresentaram o esforço que estava sendo realizado para implementação de novos sistemas de codificação de vídeo baseados no MPEG-2 e no H.264/AVC para tornar possível uma demonstração ao vivo do Super Hi-Vision. O principal objetivo era atrair a atenção pública e colecionar opiniões sobre a percepção deste sistema como um serviço de radiodifusão.

Dois aspectos importantes resultaram das demonstrações que foram viabilizadas ao longo de 2008: o primeiro aspecto, técnico, é referente à tecnologia de codificação utilizada no Super Hi-Vision para comparação com a tecnologia escalável proposta neste trabalho; o segundo aspecto, não técnico, é referente à boa impressão causada pelo sistema ao público na última edição do evento IBC (maior exibição de tecnologia em radiodifusão na Europa), evidenciando que transmissões em resoluções acima do HDTV devem ser utilizadas não apenas para transmissões especiais ou em demonstrações, mas para chegar às casas dos usuários finais. Neste sentido, grandes empresas da indústria de televisores vêm trabalhando em novos modelos com resolução 4 vezes superior ao HDTV, sendo que alguns desses modelos já se encontram em exibição.

A especificação atual do Super Hi-Vision define o formato do vídeo em 7680x4320p@60. Esse formato implica em uma quantidade de pixels 16 vezes maior que a do formato 1080p@60 e 32 vezes maior que a do formato 1080i@30. Tanto as normas de vídeo (considerando os níveis suportados atualmente) quanto a tecnologia não estão preparadas para o tratamento de toda essa informação nos processos de codificação e deco-

dificação. Na realidade, a tecnologia utilizada nas câmeras de vídeo para captura também não está preparada para essa resolução real. Dessa forma, os sistemas desenvolvidos para viabilizar as demonstrações do Super Hi-Vision realizam a segmentação da seqüência de vídeo no tempo e no espaço, permitindo o processamento de blocos de informação em paralelo na codificação e na decodificação.

Especificamente para a implementação baseada no H.264/AVC, utilizada na demonstração de transmissão entre a emissora RAI em Turin e o IBC em Amsterdam, as imagens foram segmentadas em 8 blocos espaciais e 2 temporais, aproveitando o esquema de amostragem das câmeras, que capturam os sinais vermelho, azul e verde em proporções e posições levemente distintas. Em função do esquema de amostragem, os blocos das imagens puderam ser codificados independentemente em 16 codificadores H.264/AVC no formato 1080i30 ou no formato 1080p30PsF³. Os fluxos de vídeo codificados foram reagrupados posteriormente em dois fluxos de transporte independentes para que pudessem ser transmitidos por dois canais de satélite de aproximadamente 70 Mbps de taxa.

De forma geral, resultados de experimentos realizados para a montagem da demonstração em questão evidenciaram que, para o formato de vídeo do Super Hi-Vision, já considerando a segmentação, porém com esquema de amostragem 4:2:2 e 10 bits de profundidade por amostra, seria necessária uma taxa de transmissão entre 140 e 160 Mbps (considerando a camada de transporte com áudio incluso). Esses valores mostram que implementações no estado-da-arte de codificadores H.264/AVC para esse formato já são capazes de gerar vídeo com qualidade a taxas próximas de 10 Mbps. Essa estimativa pode ser feita com boa aproximação por divisão exata (por 16 no caso), pois características estatísticas da codificação de vídeo não foram exploradas entre os codificadores nessa implementação. Em versões futuras, essas características deverão ser consideradas para melhorar o desempenho de codificação, pois a dificuldade de cada bloco não é necessariamente igual devido a variações espaciais no conteúdo na cena.

As equipes que estão trabalhando no desenvolvimento do Super Hi-Vision esperam ainda obter reduções significativas na taxa de vídeo comprimido, visando atingir 70 Mbps para transmitir vídeo em um único canal de satélite, mas para isso pode ser necessário adotar outro CODEC de vídeo. O Dirac tem sido a opção preferida como solução alternativa, embora ainda não cumpra os requisitos de eficiência com suas ferramentas atuais.

³Psf significa quadro progressivo segmentando (*Progressive segmented Frame*) e foi desenvolvido para aproveitar equipamentos com entrelaçamento para processar vídeo progressivo. A vantagem do PsF em comparação a vídeo entrelaçado nativo é que o PsF separa um quadro amostrado no mesmo instante de tempo em campos, enquanto cada campo no vídeo entrelaçado é amostrado em instante diferente.

4.1.3 Suporte à Escalabilidade no Sistema Europeu

O consórcio DVB anunciou ao final de 2008 (revista eletrônica⁴) a iniciativa de promover a extensão de suas normas para contemplar a escalabilidade de vídeo do H.264/AVC em transmissões por radiodifusão e por rede IP.

Especificamente para a radiodifusão, o objetivo da iniciativa, que não será concluída antes desta dissertação, é permitir que programas considerados *premium* possam ser recebidos pelos usuários no formato 1080p. A opção pelo SVC, neste caso, deve-se ao fato de alguns serviços baseados no DVB já terem sido lançados com a codificação H.264/AVC.

Apesar de as normas do DVB inicialmente especificarem exclusivamente o MPEG-2 como ferramenta de codificação de vídeo, a partir de certo momento elas passaram a contemplar também a codificação H.264/AVC como opção. Dessa forma, alguns países que lançaram seus serviços de transmissão digital mais recentemente adotaram essa codificação, limitada ainda ao formato 1080i. Como os investimentos envolvidos na implantação de um sistema baseado em H.264/AVC são sensivelmente superiores, a existência de receptores legados foi determinante para se optar pela escalabilidade.

Por outro lado, esta decisão não poderia ter sido tomada sem o embasamento técnico de sua viabilidade, tanto do ponto de vista lógico de compatibilidade dos novos formatos com os equipamentos legados quanto do ponto de vista de eficiência de codificação do SVC. As principais contribuições neste sentido foram organizadas e relatadas pela EBU.

Alguns relatórios estão abertos ao público, enquanto outros estão restritos ao acesso de membros dessa entidade. Os relatórios públicos (EBU, 2008; KOUADIO et al., 2008) apresentam avaliações relativas a opções de formatos de vídeo e de codificação, em função das instalações existentes, e de como ambos afetam a qualidade na percepção do usuário. Além disso, apresentam também usos potenciais do SVC em termos de aplicações e de desempenho, através de casos de uso e de evidências experimentais.

Os estudos também apontam que há uma tendência de utilização do H.264/AVC em aplicações multimídia, com ampla presença deste CODEC não somente em receptores fixos, mas de forma crescente em dispositivos móveis e portáteis. Com essa generalização, a transcodificação de conteúdo deve ser limitada em breve à adaptação de formato e taxa, sendo este um ponto chave para consideração do SVC na especificação de sistemas totalmente novos ou que sejam evoluções de sistemas em uso.

⁴DVB-SCENE: revista informativa do consórcio DVB publicada a cada 3 meses em média. A edição referenciada é a de número 28, de Dezembro de 2008. Disponível em: http://dvb.org/news_events/dvbscene_magazine/DVB-SCENE28.pdf. Acesso em: 28 jan. 2009.

De forma geral, considerando os sistemas que estavam em uso na Europa, foram identificados 4 formatos de produção de conteúdo HDTV: 720p@50, 1080p@25, 1080i@25 e 1080p@50. Observou-se que os formatos 1080i@25 e 720p@50 também estavam sendo utilizados para a radiodifusão ou outra forma de distribuição, enquanto 1080p@25 estava sendo utilizado somente para produção. No entanto, o formato 1080p@25 poderia também ser distribuído, desde que mapeado em 1080i@25 como 1080psf@25 ou convertido para 720p@50 com subamostragem espacial e repetição de quadro. Por fim, 1080p@50, que é denominado formato HDTV de “terceira geração”, ainda apresentava pouco uso, sendo considerado pelos radiodifusores o próximo formato de vídeo para produção e, possivelmente, transmissão.

Dentre os formatos citados, o 720p@50 foi apontado como o que apresenta a melhor relação entre qualidade e taxa de transmissão durante avaliações com codificadores H.264/AVC. Por qualidade considerou-se a percepção do usuário na ampla maioria dos televisores HDTV disponíveis no mercado, com varredura progressiva, resolução 1366x768 pixels e tecnologia de plasma ou LCD.

O formato 1080p50 também se mostrou eficiente em termos de qualidade, com taxas de compressão comparáveis às do formato 1080i@25. No entanto, ainda que a produção seja em 1080p@50, foi recomendada transmissão 720p@50 mesmo para televisores de até 50 polegadas, com resolução de tela de 1920x1080 pixels. O formato 1080i@25 foi considerado um formato utilizado apenas em situações específicas e normalmente por condições não técnicas, como aproveitamento de equipamento legado.

Quanto à compressão H.264/AVC sem escalabilidade, os estudos da EBU revelaram ainda uma significativa evolução de qualidade dos codificadores nos últimos anos. Os resultados de análises por PSNR, seguidas de análises subjetivas, puderam ser resumidos nas seguintes conclusões iniciais:

- A eficiência prática de codificação do H.264/AVC partiu de uma situação em que não apresentava ganho em relação à compressão MPEG-2 em 2005, sendo significativamente melhorada ao longo de 2006 e 2007.
- Algumas implementações práticas de codificadores H.264/AVC, avaliadas em 2008, já permitiam economia de taxa de transmissão entre 40 e 50% (dependendo da dificuldade do conteúdo) se comparada ao MPEG-2.
- É geralmente mais difícil comprimir 1080i@25 do que 720p@50, com uma vantagem de economia em torno de 20% com o segundo formato dependendo do conteúdo.

Adicionalmente, quanto à compressão H.264/AVC com escalabilidade SVC, os estudos da EBU revelaram que, partindo de uma camada base, seja em 720p@50 ou em 1080i@25, para uma camada de enriquecimento de 1080p@50, os resultados são semelhantes em valores finais de PSNR (diferença inferior a 0,5 dB) para taxas finais iguais. Já comparados ao *simulcast*, ambos os casos apresentaram ganho de taxa em torno de 30%.

Os testes foram realizados com equipamentos e softwares dotados de implementações no estado-da-arte, no entanto apontou-se que avaliações visuais deveriam ser refeitas quando codificadores com novas otimizações estiverem disponíveis, pois havia claramente espaço para melhorias nas implementações. Naquele momento, os resultados experimentais mostraram que uma camada simples H.264/AVC em 1080p iria requerer em torno 13 Mbps para uma qualidade razoável. A adição de um sinal 720p, para atender ao legado de receptores, iria requerer 8 Mbps adicionais, somando um total de 21 Mbps. Alternativamente, 15,4 Mbps foi encontrado como taxa necessária para uma transmissão SVC com duas camadas: uma em 720p para o legado e outra de enriquecimento para 1080p. Foram avaliadas também outras resoluções não suportadas no SBTVD como 1440x1080 pixels.

4.2 Considerações sobre o Sistema Brasileiro

Dos projetos correlatos apresentados anteriormente, a análise do JVT marca os primeiros resultados efetivos de utilização do SVC, especialmente para a radiodifusão, já sinalizando que esta poderia ser uma ferramenta de evolução de sistemas baseados em H.264/AVC, como é o caso do SBTVD. O Super Hi-Vision demonstra a expectativa de longo prazo por resoluções acima do HDTV e também os esforços realizados para contornar limitações tecnológicas e mesmo de especificações para viabilizar sistemas avançados. Este projeto evidenciou experimentalmente o conceito de transmissão do mesmo fluxo de vídeo dividido entre dois canais simultâneos, embora com o artifício de segmentação dos quadros de vídeo e não de escalabilidade, o que pode ser estendido e adaptado para o SBTVD. Já o projeto europeu, equivalente direto ao projeto proposto nesta dissertação, tornou-se público ao longo de 2008, atualizando as análises subjetivas e objetivas de transmissões baseadas em H.264/AVC sem e com escalabilidade, sistematizando o processo de testes, análise e evolução do sistema considerando toda sua cadeia.

No processo de avaliação do sistema europeu, no entanto, pelo menos três pontos, que podem ter reflexos importantes nos resultados, devem ser observados para analogia com o sistema brasileiro: a base da taxa de quadros; a evolução dos codificadores para vídeo entrelaçado; e a prioridade por multi-programação.

Quanto à taxa de quadros, a base dos sistemas implantados na Europa é de 25 ou 50 quadros por segundo, enquanto no Brasil, a taxa é de 29,97 ou 59,94. Desta forma, espera-se que os resultados evidenciados nos experimentos apresentem valores de relação entre taxa de bits e qualidade de vídeo um pouco inferiores para seqüências produzidas com as características brasileiras. Espera-se, no entanto, que a queda de desempenho não seja proporcional à diferença entre as taxas de quadros, pois com o aumento da taxa de quadros aumenta também a correlação temporal entre imagens consecutivas, o que contribui, em contra-partida, para melhor eficiência da codificação.

A questão do entrelaçamento também pode ter influência em resultados, pois na Europa, sensivelmente, a maior parte da produção HDTV é transmitida em 720p e não em 1080i, como apontam os relatórios. Dessa forma, é natural que as empresas tenham investido mais esforço de otimização na implementação das ferramentas de escalabilidade para vídeo progressivo. Já no Brasil, é possível observar varrendo as emissoras em um televisor com receptor digital, que toda a programação HDTV, pelo menos na cidade de São Paulo, está sendo transmitida no formato 1080i. Pode haver, portanto, espaço para otimizações mais significativas nos algoritmos de compressão para escalabilidade a partir de vídeo entrelaçado, alterando também o quadro de resultados apontados pela EBU.

Na Europa, as transmissões digitais têm priorizado a multi-programação, o que demanda acomodação do espectro de radiofrequências de forma mais racionada. Essa característica certamente deve ter sido levada em consideração nas relações de compromisso entre taxa e qualidade para consolidar as recomendações sobre formatos de transmissão. No Brasil, pelo simples método de varredura dos canais digitais descrito anteriormente, observa-se que as emissoras, novamente pelo menos na cidade de São Paulo, não estão investindo em multiprogramação, mas sim na alta resolução. Dessa forma, o espectro já destinado às emissoras apresenta maior potencial de exploração.

Por essas e outras particularidades do sistema brasileiro, seria importante que o Fórum SBTVD realizasse ou incentivasse o mapeamento das características objetivas e subjetivas das transmissões digitais, considerando toda a cadeia, desde a produção até a recepção. Com isso seria possível gerar relatórios técnicos similares aos da EBU, com uma análise criteriosa para as tomadas de decisão sobre a evolução dos formatos de transmissão no SBTVD.

Paralelamente, no entanto, pode-se definir alguns casos de uso desejados para direcionar as avaliações e a montagem de testes, o que será apresentado adiante com base em características atuais do SBTVD e possibilidades vislumbradas com a escalabilidade.

4.2.1 Receptores

Todos os receptores compatíveis com o SBTVD devem contemplar a decodificação de vídeo H.264/AVC em perfil HIGH e nível igual ou superior ao 4.0. Observando a Tabela 2.3, conclui-se que todo receptor legado deve ser capaz de decodificar vídeo em formato HDTV, mesmo que o televisor onde o vídeo está sendo apresentado seja de resolução SDTV.

Dessa forma, já não há necessidade no sistema atual de se transmitir o mesmo conteúdo nestes dois formatos simultaneamente. Nos cenários futuros, se uma determinada emissora optar por transmitir vídeo em novos formatos para receptores fixos, superiores aos atualmente suportados, sendo a transmissão escalável mais econômica que o *simulcast*, também não haverá necessidade de transmissão do conteúdo de alta resolução simultaneamente em formato 480i ou 480p.

Essa constatação permite assumir que qualquer transmissão com escalabilidade visando novos formatos deverá partir de uma camada base codificada já em um formato HDTV. Embora pelos parâmetros permitidos na norma de codificação de vídeo, os formatos 720p@29,97 e 1080p@29,97^{5 6} devessem ser suportados pelo sistema, eles não são recomendados nos guias operacionais e a decodificação nos receptores, apesar de provavelmente possível pelas características de implementação dos decodificadores, pode não ser garantida. Portanto, os formatos que devem ser considerados inicialmente como camada base são 720p@60 ou 1080i@30.

Quanto aos perfis de codificação, hoje o sistema suporta transmissão para receptores fixos tanto no perfil MAIN quanto no perfil HIGH. Para transmissões com escalabilidade qualquer desses perfis pode continuar sendo utilizado como camada base para compatibilidade com a camada de enriquecimento, necessariamente codificada com o perfil Scalable HIGH.

4.2.2 Particularidades dos Formatos de Vídeo

Com relação ao tipo de varredura, o SVC permite a codificação de vídeo entre camadas com qualquer combinação: progressiva-progressiva; progressiva-entrelaçada; entrelaçada-

⁵Juntamente com o Prof. Dr. Marcelo K. Zuffo, o autor defendeu que as combinações de formatos e taxas de quadros não fosse tão restrita no SBTVD, dado que os decodificadores, pela forma de controle de quadros do H.264/AVC, são flexíveis. No entanto, principalmente por contraposição da indústria de receptores, as combinações acabaram sendo reduzidas.

⁶Deste ponto em diante para simplificar o texto, as frequências de quadros do SBTVD serão arredondadas. Dessa forma, 720p@59,94 será redigido como 720p@60 e 1080i@29,97 será redigido como 1080i@30. O mesmo vale para 2160p.

progressiva; e entrelaçada-entrelaçada. No entanto, como as ferramentas de codificação para entrelaçamento não são suportadas a partir do nível 4.2 (inclusive), o único caso de interesse para o SBTVD é o entrelaçado-progressivo, quando a camada base consiste em vídeo no formato 1080i. Neste caso, o SVC permite duas configurações distintas na camada base:

- Camada base codificada em quadros, utilizando a ferramenta MBAFF.
- Camada base codificada em campos pertencentes a unidades de acesso distintas.

François, Viéron e Bottreau (2007) descrevem com detalhes as características da codificação escalável com entrelaçamento no SVC, apresentando o impacto de codificação em função do tipo de camada base selecionada. A escolha da camada base afeta principalmente a eficiência da codificação.

A partir de 1080i@30, que é o formato de interesse para o SBTVD, com os níveis atualmente definidos no SVC, seriam suportados os formatos 1080p@60 (nível 4.2), 1080p@120 (nível 5.1) e 2160p@30 (nível 5.1) (considerando apenas formatos com relação de aspecto 16x9).

No entanto, de acordo com as restrições de parâmetros dos elementos sintáticos de um fluxo codificado em SVC, a única possibilidade de transmissão de vídeo 2160p@30 no SBTVD seria a partir de uma camada base em 1080i@30 codificada em quadros com a ferramenta MBAFF ativa. Neste caso, no entanto, não seria possível a transmissão de uma camada intermediária em 1080p@60 ou em 1080p@120, pois a norma não permite fluxos com diminuição da taxa de quadros conforme aumenta a camada de enriquecimento.

Esse ponto de decisão é importante, pois se existe a possibilidade de transmissão de conteúdo produzido em 2160p@30, a próxima geração de receptores deveria ser especificada já com suporte a este formato. Uma vez que se especifiquem novos receptores com suporte somente a 1080p@60, serão necessários novos níveis do SVC para permitir a codificação de vídeo em 2160p@60 para uma terceira geração de receptores (até este ponto não estão sendo consideradas para análise e argumentação as taxas necessárias para transmissão de cada formato).

Vale, por fim, observar que a relação entre as resoluções 720p e 1080p não é inteira ($1080/720 = 1,5$), o que não é permitido em alguns perfis de escalabilidade. O perfil Scalable HIGH, no entanto, oferece sinalização de escalabilidade estendida, que permite relações arbitrárias entre as resoluções das camadas codificadas.

4.2.3 Possibilidades com a Escalabilidade

Apesar de toda a flexibilidade conceitual em que está apoiado o SVC, limitações de seus parâmetros em função principalmente das tecnologias disponíveis para processamento, transmissão e armazenamento, impõem restrições temporárias aos cenários de evolução da transmissão digital como discutido anteriormente. Estas restrições devem ser consideradas para a evolução de curto prazo do sistema, mas serão desprezadas para considerações adicionais de longo prazo.

Como resumo do cenário atual, assumindo o SVC como ferramenta natural para a escalabilidade de vídeo no SBTVD, necessariamente deve ser utilizado o perfil Scalable HIGH em função de compatibilidade com o sistema legado. Além disso, é sensato assumir que a camada base de uma transmissão escalável deva ser configurada em 720p@60 ou 1080i@30, como justificado na Seção 4.2.1.

Com este cenário de partida, é necessário estabelecer o nível de codificação mais adequado a ser suportado pela próxima geração de transmissores e receptores. Neste sentido, embora o primeiro pensamento seja o de garantir maior flexibilidade ao sistema, deve-se ter cautela quanto à disponibilidade de pastilhas (*chips*) e equipamentos nos próximos anos e quanto à disponibilidade de taxa nos canais digitais.

Durante os trabalhos realizados na segunda fase do SBTVD em 2005, quando a extensão de fidelidade acabara de ser concluída, as pastilhas capazes de decodificar H.264/AVC no perfil MAIN e nível 4.0 ainda estavam sendo corrigidas e finalizadas. Somente em 2006 surgiram pastilhas capazes de decodificar o perfil HIGH, inicialmente no nível 4.0 e ainda com poucas opções de fabricantes no mercado. Em 2007, durante os trabalhos de especificação das normas no Fórum SBTVD, investigou-se a disponibilidade de pastilhas que decodificassem o perfil HIGH no nível 4.2, que já permitiria suporte ao formato 1080p@60 no SBTVD. Chegou-se à conclusão que esta opção seria arriscada do ponto de vista de disponibilidade e de custos.

Atualmente, são encontradas diversas opções de pastilhas que decodificam perfil HIGH no nível 4.0, principalmente voltadas para radiodifusão, e também que decodificam nível 4.1, para discos Blu-RayTM. Algumas soluções estão surgindo para o nível 4.2, voltadas para serviços de vídeo por satélite, mas ainda são raras. Todas estas opções, no entanto, não oferecem suporte à escalabilidade.

Com este panorama, é de se esperar que nos próximos 2 ou 3 anos as implementações no nível 4.2, tanto do perfil HIGH quanto do perfil Scalable HIGH, sejam consolidadas.

O nível 5.1, que seria outro nível de interesse para o SBTVD, representa um aumento em torno de 4 vezes nos requisitos de processamento e memória (derivado da Tabela 2.3), apontando que este perfil deve tornar-se disponível apenas em um cenário de longo prazo.

Quanto à disponibilidade de taxa de um canal digital, será considerada uma configuração hipotética do SBTVD para uma análise inicial. Nesta configuração hipotética, um segmento é destinado à transmissão portátil e os outros 12 segmentos à transmissão de um único programa. Além disso, os 12 segmentos de interesse estão configurados com modulação 64-QAM. Nestas condições, dependendo de alguns outros parâmetros da modulação, a taxa de bits disponível para o programa principal varia entre 10,1 Mbps e 21,4 Mbps, sendo que esta taxa deve ser compartilhada entre áudio, vídeo e dados.

Supondo que o vídeo neste sistema consuma entre 80 e 90% da taxa de transmissão, pelos resultados apresentados dos projetos correlatos, conclui-se que este canal pode vir a contemplar a transmissão de um vídeo escalável codificado com uma camada de enriquecimento no formato máximo de 1080p@60. E, ainda assim, para viabilizar esse tipo de transmissão, aparentemente será necessário esforço de otimização nos codificadores.

Caso o acompanhamento da evolução de eficiência dos codificadores mostre que há uma assíntota de saturação, na qual a relação entre a taxa disponível e a qualidade possível de se obter com a escalabilidade ainda seja insuficiente, como alternativa pode-se investigar a transmissão da camada de enriquecimento em uma camada hierárquica de modulação distinta.

Dessa forma, cada camada hierárquica poderia ser configurada independentemente, permitindo maior controle sobre a transmissão, otimizando para cada uma o compromisso entre taxa total, potência de transmissão, robustez e alcance do sinal. Para isso, no entanto, seriam necessárias modificações às normas de multiplexação e transmissão, pois ambas não prevêem a separação de um fluxo elementar em fluxos de transporte distintos. Neste sistema, a princípio, a escalabilidade por qualidade não teria aplicação.

Pensando em cenários futuros, a evolução das normas de multiplexação e transmissão poderia permitir inclusive a separação das camadas de enriquecimento em canais distintos. Os canais para transmissão do conteúdo adicional não necessariamente seriam novas concessões para as emissoras que já operam um canal digital. Além de uma emissora poder comercializar sua taxa que eventualmente esteja sem uso, poderiam haver novos prestadores de serviços destinados a alocar dinamicamente fatias de sua transmissão para emissoras interessadas em transmitir algum programa especial utilizando mais camadas de enriquecimento. As principais motivações para esse modelo estão no bom uso do es-

pectro e no fato de as emissoras estarem investindo em equipamentos para produção em 1080i@30 para o sistema recém lançado. Dificilmente estas emissoras teriam condições de reinvestir para substituir todos os seus equipamentos novamente. Dessa forma, a oferta de programação especial seria gradual, até que novos serviços, como o próprio Super Hi-Vision estejam disponíveis.

4.3 Ambiente de Investigação Experimental

Com base nas considerações apresentadas sobre o sistema brasileiro e sobre as possibilidades de evolução pela escalabilidade de vídeo, esta seção apresenta o ambiente especificado e montado para realizar tanto algumas investigações iniciais, cujos resultados são apresentados no Capítulo 5, quanto investigações futuras.

4.3.1 Requisitos

Como linha geral, os requisitos definidos para a montagem do ambiente experimental visaram possibilitar investigações sobre a situação atual da radiodifusão digital terrestre e aberta e sobre algumas premissas utilizadas com base em outros trabalhos citados no texto. Com estes resultados, pretende-se estabelecer referências próprias para a realização de trabalhos futuros, que também devem ser contemplados pelo ambiente proposto, no sentido de acompanhar a evolução das ferramentas de codificação e de realizar ensaios de evolução do SBTVD.

Dessa forma, o ambiente experimental deste trabalho deveria ser capaz de permitir:

- Análise do espectro atualmente reservado para a transmissão digital, para o mapeamento da ocupação atual do sistema e do potencial de utilização com novos serviços e soluções.
- Análise das características dos programas que estão sendo transmitidos pelas radiodifusoras, principalmente no que diz respeito à taxa útil⁷ de cada canal para o programa principal, em especial para o fluxo de vídeo.
- Codificação flexível de vídeo H.264/AVC e de vídeo SVC.

⁷Taxa útil: refere-se ao valor máximo permitido para a transmissão de dados na camada principal da modulação OFDM configurada em cada canal. Este valor máximo é derivado de alguns parâmetros da modulação em consulta à respectiva norma do SBTVD.

- Multiplexação de um fluxo elementar de vídeo H.264/AVC, em qualquer das configurações anteriores, em um fluxo de bits tipo TS compatível com o SBTVD.
- Re-multiplexação do TS gerado para transmissão em modulação BST-OFDM também em conformidade com o SBTVD.
- Recepção de sinais digitais abertos ou produzidos no próprio ambiente experimental.
- Análises objetivas de qualidade de vídeo e inspeção visual nas imagens de seqüências de vídeo codificadas e não codificadas⁸.

Diversos equipamentos e softwares foram analisados ao longo do trabalho para atender aos requisitos especificados. As maiores dificuldades encontradas na montagem do ambiente de investigação foram em relação à compatibilidade de equipamentos com o SBTVD e em relação à inspeção visual de vídeo.

4.3.2 Sintonizadores Digitais

Dois receptores de fabricantes diferentes, disponíveis em kits de avaliação, foram considerados no ambiente de investigação para recepção e análise da camada física do sinal digital dos canais em operação no SBTVD.

O primeiro receptor, composto por sintonizador e demodulador da Fujitsu, denominado *Fujitsu MB86A20S Evaluation Board*, possui interface de controle e comunicação por USB. Em um computador x86 pode ser rodado um software de controle e análise, cuja tela principal é apresentada na Figura 4.1.

Em vermelho estão indicados os principais parâmetros de configuração de um canal digital a partir dos quais é possível determinar a taxa útil disponível para a camada lógica do TS e conseqüentemente para o fluxo de vídeo no canal. À esquerda, está circulado o campo que identifica o canal em análise, seguido pela freqüência do canal à direita. Estes campos não têm influência sobre a taxa útil da transmissão, mas serão utilizados como referência nas tabelas de resultados. Ao lado direito, de cima para baixo, podem ser observados em seqüência os campos que identificam o intervalo de guarda, o tipo de constelação utilizado na modulação das portadoras, a taxa do código convolucional interno e a quantidade de segmentos utilizados na camada da programação principal.

⁸Num primeiro momento, inspeção visual refere-se apenas à visualização de imagens decodificadas, não necessariamente na taxa e na resolução correta do vídeo. A montagem de um ambiente para avaliações subjetivas não fez parte do escopo deste trabalho e nem, portanto, dos requisitos considerados

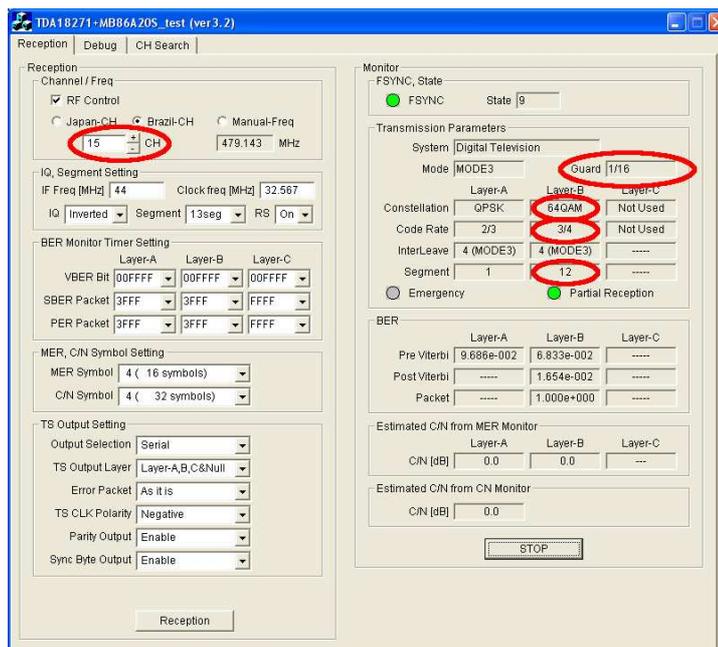


Figura 4.1: Tela de análise de um canal COFDM do software da Fujitsu

Com base nos três primeiros parâmetros, consultando-se a norma de modulação do SBTVD (ABNT, 2007a), determina-se a taxa útil por segmento da camada principal. A taxa útil total da camada é então obtida pelo produto entre a taxa por segmento e a quantidade de segmentos.

A limitação encontrada neste receptor, para os requisitos deste trabalho, foi a indisponibilidade de uma interface padrão DVB-SPI, utilizada em muitos equipamentos de medição e análise de TS para aquisição de conteúdo. Desta forma, não foi possível a integração deste receptor para análise das camadas lógicas de transporte e das estruturas e características do fluxo de vídeo.

O segundo receptor avaliado e utilizado neste trabalho foi o *Integrated Tuner Demodulator - TDH Evaluation Board*, da Alps Electric. Este, por sua vez, apesar de não apresentar uma aplicação rica de controle e análise (pelo menos na versão recebida para testes), continha em sua placa a interface DVB-SPI necessária para as análises lógicas.

4.3.3 Analisadores de Fluxos de Transporte e Vídeo

Equipamentos e aplicativos para análise da camada lógica dos fluxos de transporte (TS) e, principalmente, do fluxo elementar de vídeo (ES) costumam ser produtos caros e de difícil acesso. Para o ambiente experimental foi possível dispor do equipamento RTX100B e do pacote de software MTS400EA *TS Compliance Analyser*, ambos da Tektronix. Estes

materiais haviam sido adquiridos em projetos anteriores relacionados à televisão digital e apresentam ferramentas completas tanto de análise quanto de transmissão, embora não estejam preparadas para as características particulares de um fluxo elementar de vídeo codificado com escalabilidade.

Para análise, o equipamento RTX100B é capaz de realizar a aquisição de um trecho de fluxo de transporte através de interface DVB-SPI, armazenando o conteúdo em disco. Esse conteúdo pode, então, ser processado pelos aplicativos de análise do MTS400EA.

Para transmissão de sinal digital, existe um aplicativo de multiplexação de sinal no MTS400EA que encapsula um fluxo elementar de vídeo em uma camada de transporte com as tabelas de informação necessárias. Outro software do próprio RTX100B permite a re-multiplexação do TS, adicionando as informações de modulação necessárias. O TS modificado para atender as características de modulação hierárquica da norma pode ser transferido para o RTX100B, que por sua vez é capaz de realizar a transmissão do TS em sinal digital BST-OFDM através de conector tipo F para cabo coaxial.

A Figura 4.2 apresenta duas telas de exemplo do software *MPEG TS Compliance*

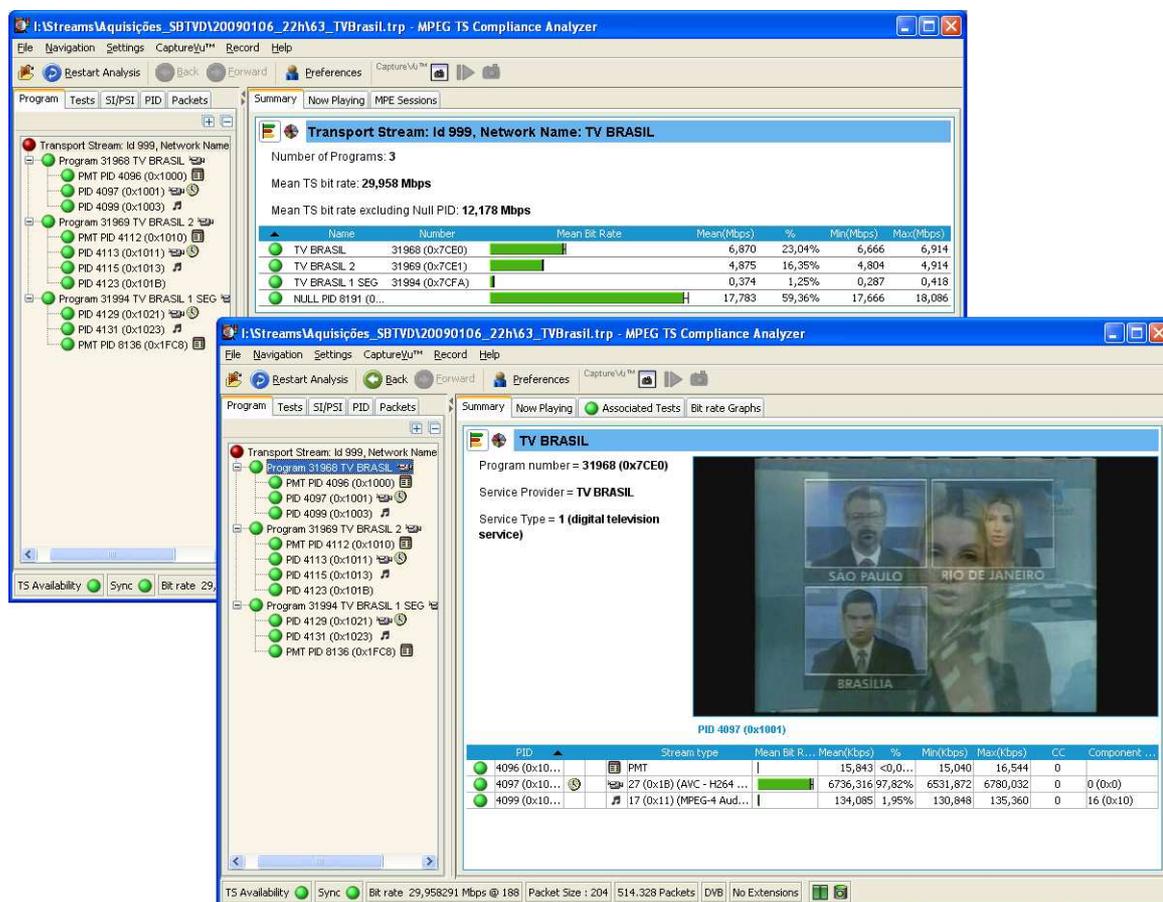


Figura 4.2: Telas de análise com software MTS400 da Tektronix

Analyser, que faz parte do pacote MTS400EA e pode ser utilizado para levantamento das características dos fluxos de bits que estão sendo transmitidos pelos radiodifusores e também de fluxos de bits gerados a partir de testes de codificação.

Muitas funcionalidades dos materiais da Tektronix, relacionadas, por exemplo, a análises de conformidade e multiplexação de tabelas, estão implementadas em roteiros (*scripts*) que podem ser modificados para implementar novas características no sistema digital para testes.

4.3.4 Software de Referência do SVC

O software de referência do projeto SVC é o *Joint Scalable Video Model* (JSVM), que foi desenvolvido principalmente por membros do grupo JVT em linguagem C++. O software é distribuído como código-fonte a partir de um servidor em Aachen, na Alemanha.

Para se obter a versão 9.15, que é a versão utilizada neste trabalho, é necessário ter instalado um cliente CVS. A versão pode ser obtida do servidor com os seguintes comandos:

```
1 cvs -d :pserver:jvtuser:jvt.Amd.2@garcon.ient.rwth-aachen.de:/cvs/jvt login
2 cvs -d :pserver:jvtuser@garcon.ient.rwth-aachen.de:/cvs/jvt co -r JSVM_9_15 jsvm
```

Após baixar o código-fonte do servidor é necessário compilar os aplicativos que compõem o software. A compilação pode ser realizada em diferentes sistemas operacionais, incluindo distribuições Linux e o Microsoft WindowsTM. Para este trabalho foram realizadas compilações nos dois sistemas. No ambiente Linux (distribuição Ubuntu 8.10), foram utilizadas ferramentas livres que podem ser instaladas pelo gerenciador de pacotes da distribuição. No ambiente Microsoft, foi necessário instalar o pacote Microsoft Visual StudioTM versão 8.0.

Junto com o código-fonte do JSVM, na raiz do projeto, está disponível o manual de software referente à versão baixada. O manual apresenta informações gerais de como baixar, compilar e executar cada aplicativo que compõe o pacote, exemplos de uso em forma de tutoriais e informações para quem deseja integrar funcionalidades novas.

Segue uma visão geral dos aplicativos que formam o JSVM com suas respectivas funcionalidades:

DownConvertStatic: permite gerar versões de uma seqüência original com resoluções espaço-temporais diferentes, para que sejam utilizadas na codificação escalável re-

presentando camadas de resolução distintas. O aplicativo de codificação não realiza esta etapa automaticamente, sendo necessário gerar todas as camadas com este aplicativo, ou similar, antes de uma codificação.

H264AVCEncoderLibTestStatic: pode ser usado para gerar fluxos de vídeo codificados em conformidade com o H.264/AVC sem escalabilidade ou com o SVC. Deve ser passado para o aplicativo, em cada chamada para execução, pelo menos um arquivo contendo a configuração dos parâmetros de codificação. Este arquivo, no caso de codificação SVC deve apontar para outros arquivos, cada um descrevendo uma camada de codificação distinta. Parâmetros de linha de comando podem ser passados ao aplicativo para sobrescrever valores dos arquivos de configuração. A eficiência da codificação depende da combinação de valores, não sendo garantida pela aplicação.

H264AVCDecoderLibTestStatic: assim como o codificador, pode ser usado para decodificar um fluxo de vídeo H.264/AVC ou SVC, gerando uma seqüência de saída sem compressão, referente à maior camada de enriquecimento. Para decodificar uma camada intermediária, é necessário utilizar o aplicativo *BitStreamExtractorStatic* antes.

AvcRewriterStatic: converte um fluxo de bits SVC em um fluxo H.264/AVC sem escalabilidade. Para isso é necessário que o fluxo SVC tenha sido especialmente codificado com suporte a esta funcionalidade (identificada do fluxo através da marcação *avc_rewrite_flag*).

BitStreamExtractorStatic: pode ser usado para extrair sub-fluxos com menor resolução espaço-temporal ou com qualidade de um fluxo H.264/AVC ou SVC original. Os sub-fluxos extraídos mantêm compatibilidade com as especificações do H.264/AVC ou do SVC.

QualityLevelAssignerStatic: pode ser usado para acrescentar informações sobre camadas de qualidade em um fluxo SVC. Esta informação, que pode ser inserida tanto no cabeçalho estendido das unidades NAL quanto em mensagens do tipo SEI, pode ser utilizada pelo processo de extração de sub-fluxos do fluxo principal.

MCTFPreProcessorStatic: pode ser usado para uma pré-filtragem da seqüência de vídeo antes da codificação. Para algumas seqüências de vídeo, este aplicativo melhora a eficiência de codificação. Inicialmente a ferramenta MCTF, na qual está baseado o aplicativo, fazia parte do fluxo principal de codificação. Quando essa

ferramenta foi excluída da norma, ela foi transformada neste aplicativo de pré-processamento.

PSNRStatic: pode ser usado para medir o PSNR entre duas seqüências de vídeo. Adicionalmente, pode também ser utilizado para calcular taxa de bits de fluxos codificados.

FixedQPEncoderStatic: pode ser usado para encontrar o valor base do coeficiente de quantização a ser utilizado na codificação de uma seqüência de vídeo a fim de se obter um determinado valor de taxa de bits ou de qualidade. Serve tanto para codificações em camada simples quanto para camadas específicas de uma codificação escalável.

SIPAnalyser: pode ser usado para fazer a decisão do modo de predição seletiva entre camadas para a codificação escalável. Este aplicativo necessita de um arquivo contendo as informações de bits por quadro, gerado pelo aplicativo de codificação com a opção “-anasip” habilitada.

Especificamente sobre o aplicativo de codificação (*H264AVCEncoderLibTestStatic*), algumas características adicionais são destacadas.

Durante a codificação, os arquivos reconstruídos de cada camada são gerados automaticamente, pois suas respectivas imagens são necessárias para as ferramentas de predição e compensação de movimento. No entanto, estes arquivos ocupam muito espaço em disco, pois são seqüências não comprimidas. Para a maior parte dos roteiros de testes realizados, não vale a pena manter estes arquivos em disco, mesmo porque eles podem ser gerados posteriormente através da decodificação, que é um processo muito mais rápido do que a codificação.

Durante a codificação, é gerado na saída padrão da janela onde o aplicativo está sendo executado um relatório apresentando características de cada imagem codificada, detalhando o tipo (I, P, B), a qualidade PSNR e a quantidade de bits por imagem. Ao final são apresentadas as médias de qualidade e de taxa de bits por camada.

O aplicativo de codificação possui dois modos de operação:

- Modo AVC legado (*AVCMode = 1*): permite a codificação de uma camada simples, sem suporte a qualquer parâmetro ou ferramenta especificado na extensão de escalabilidade da norma H.264/AVC. Neste modo, não há qualquer controle de taxa, de forma que a codificação ocorre empregando o mesmo coeficiente de quantização em

todos os macroblocos. Por outro lado, neste modo é possível definir a estrutura da seqüência de codificação livremente, respeitando apenas as limitações do perfil e do nível escolhidos para a codificação.

- Modo SVC ($AVCMode = 0$): codifica a seqüência de vídeo de acordo com a extensão de escalabilidade do H.264/AVC, permitindo a codificação de um fluxo em camada simples ou em múltiplas camadas. Neste modo, pode ser habilitado controle de taxa para a camada base durante a codificação. No entanto, este controle não se aplica a quadros do tipo B e nem a vídeos entrelaçados, sendo ainda uma implementação bastante limitada. A estrutura utilizada para codificação neste modo é a estrutura hierárquica de quadros tipo B, não sendo possível, portanto, a configuração livre da seqüência de quadros.

Como as normas de codificação de vídeo não definem os algoritmos que devem ser utilizados pelos codificadores, como visto no Capítulo 2, o JSVM utiliza equações heurísticas calibradas por levantamentos estatísticos com seqüências de vídeo para implementar funções de decisão. Dessa forma, pequenas variações sobre alguns parâmetros de codificação podem gerar diferenças expressivas na eficiência do fluxo gerado. Na codificação hierárquica, por exemplo, os coeficientes de quantização de cada camada temporal, por padrão, são derivados do coeficiente especificado como referência no arquivo de configuração. Entretanto, o coeficiente de quantização especificado também é utilizado por alguns algoritmos de decisão. Dessa forma, se o coeficiente de referência for alterado, mesmo forçando os coeficientes para que em cada camada eles assumam os valores que assumiriam naturalmente na situação anterior, os valores passados para as equações de decisão não são os mesmos, o que pode gerar diferenças significativas de eficiência.

4.3.5 Outros Equipamentos e Softwares

Além dos materiais já apresentados em detalhes, alguns outros que foram avaliados para compor o ambiente de investigação experimental estão relacionados a seguir:

ImageMagick: é uma ferramenta para conversão entre formatos de imagens e de seqüências de vídeo sem compressão. Essa ferramenta está disponível gratuitamente⁹ e possui compilações para ambiente Linux e Microsoft WindowsTM. Ela se aplica, por exemplo, à conversão de seqüências no formato sgi16¹⁰, que é o caso das seqüências

⁹Disponível em: <http://www.imagemagick.org/script/index.php>. Acesso em: set. 2008.

¹⁰Especificação disponível em: ftp://vqeg.its.blrdoc.gov/HDTV/SVT_MultiFormat/. Acesso em: set. 2008.

que serão apresentadas adiante, em formato I420¹¹, que é o formato esperado pelos aplicativos do JSVM.

MPlayer/MEncoder: estes dois aplicativos fazem parte do pacote MPlayer, que é distribuído como software livre¹² em código-fonte ou em binários pré-compilados para execução tanto em Linux quanto em Microsoft WindowsTM. O MPlayer é capaz de reproduzir seqüências de vídeo comprimido em diversos formatos, incluindo H.264/AVC sem escalabilidade, e também seqüências sem compressão. O MEncoder, por sua vez, permite a compressão de seqüências de vídeos também em diversos formatos ou simplesmente o encapsulamento de um fluxo elementar de vídeo em uma camada de transporte.

Kyrion File Encoder: é uma ferramenta comercial¹³ de codificação de vídeo da empresa francesa AteMe, capaz de codificar vídeo H.264/AVC sem escalabilidade, porém com suporte à codificação hierárquica por quadros tipo B. A ferramenta gera vídeo já encapsulado em uma camada de transporte MPEG-2, pronta para ser re-multiplexada para transmissão. No entanto, o software exige que o vídeo de entrada para codificação esteja também encapsulado em uma camada de transporte, ou seja, a ferramenta não aceita um fluxo elementar de vídeo como entrada. Além disso, não há muita flexibilidade para codificar vídeo entrelaçado e os parâmetros de temporização, restritos a valores específicos pela norma do SBTVD, não podem ser especificados na ferramenta, o que faz com que seja produzido um fluxo nem sempre inteiramente compatível com o SBTVD.

SVC Analyzer: é uma ferramenta comercial¹⁴ exclusivamente para plataforma Microsoft WindowsTM destinada à análise de fluxos de vídeo compatíveis com a atual especificação da norma H.264/AVC, incluindo sua extensão de escalabilidade. Com esta ferramenta é possível escolher a camada a ser decodificada para análise, bem como ver detalhes das unidades de acesso e das unidades NAL que compõem o fluxo de vídeo. É possível também visualizar detalhes de codificação de imagens, de fatias e de macroblocos, incluindo coeficientes, vetores de movimento, etc.

AVISynth: é uma ferramenta distribuída como software livre¹⁵ destinada à pós-produção

¹¹Esse tipo de classificação de formato de vídeo por quatro caracteres é conhecido como *Four Characters Code* (FourCC) e uma lista atualizada com os formatos conhecidos fica disponível em: <http://www.fourcc.org/>. Acesso em: set. 2008.

¹²Disponível em: <http://www.mplayerhq.hu/design7/news.html>. Acesso em: set. 2008.

¹³Disponível em: <http://www.ateme.com/bb/>. Acesso em: set. 2008.

¹⁴Disponível em: <http://www.svc-analyzer.com/>. Acesso em: set. 2008.

¹⁵Disponível em: http://avisynth.org/mediawiki/Main_Page. Acesso em: set. 2008.

de vídeo. Na verdade, a ferramenta, que não apresenta interface gráfica, fornece um poderoso ambiente para tratamento de vídeo através de roteiros (*scripts*) que interligam módulos de processamento (*plug-ins*) para edição em tempo real. Dentre os módulos disponíveis para este ambiente, encontram-se módulos para filtragem, escalonamento, medidas objetivas (PSNR e SSIM, por exemplo) e visualização. Dos módulos adicionais disponíveis, neste trabalho foram necessários o RawSource e o SSIM.

Ffdshow-tryouts: é um conjunto de CODECs¹⁶ com suporte a uma vasta gama de formatos de áudio e vídeo. Inclui diversos tipos de filtros que se integram à arquitetura DirectShow do Microsoft WindowsTM, disponibilizando suas funcionalidades para aplicativos como o AVISynth, por exemplo.

Cygwin: é uma ferramenta que não está relacionada diretamente a vídeo. Serve para simular um ambiente Linux em ambiente Microsoft WindowsTM, desta forma, permite que roteiros tipo *bash* sejam rodados para automatizar tarefas. Esta ferramenta também é distribuída como software livre¹⁷.

DTA-122-SX: é uma placa de aquisição¹⁸ para computador com entrada DVB-SPI. Pertence ao fabricante dektec e pode ser utilizada para fazer aquisições de fluxos de transporte de canais sintonizados em um receptor externo com saída DVB-SPI.

Receptor SBTVD: é um receptor de TV Digital para o SBTVD, desenvolvido pela equipe do LSI com base no SoC CE2110 da Intel, com saídas de vídeo composto e HDMI. Pode ser utilizado para validar a compatibilidade de um sinal gerado a partir de um fluxo de vídeo com escalabilidade para evidenciar a manutenção dos receptores legados.

4.3.6 Seqüências de Vídeo

Para complementar o ambiente de investigação, é fundamental a disponibilidade de seqüências de vídeo com alta qualidade, de preferência em seus formatos originais (sem ter passado por processos que envolvam perdas). Esse tipo de material é relativamente fácil de ser encontrado em formatos de baixa resolução, disponíveis sem custos na internet. Por outro lado, material de alta resolução e além, normalmente são comercializados, envolvendo altos custos.

¹⁶Disponível em: <http://ffdshow-tryout.sourceforge.net/>. Acesso em: set. 2008.

¹⁷Disponível em: www.cygwin.com. Acesso em: set. 2008.

¹⁸Disponível em: <http://www.dektec.com/>. Acesso em: set. 2008.

Por este motivo, para o início das investigações neste trabalho, foram baixados materiais¹⁹ produzidos pela radiodifusora sueca *Sveriges Television* (SVT) para fins de pesquisa e desenvolvimento.

O conteúdo original foi capturado em 3840x2160p@50 e digitalizado no formato sgi16, que representa cada imagem das seqüências de vídeo em um arquivo separado. A partir do formato original 3840x2160p@50, a emissora derivou outros formatos utilizando ferramentas de subamostragem conforme descrito no documento que acompanha o material:

- 1920x1080p@50
- 1920x1080i@25
- 1280x720p@50
- 720x576i@25

Existem 5 seqüências disponíveis, cujas primeiras imagens são apresentadas na Figura 4.3.

Segundo avaliação da emissora, cada seqüência apresenta uma dificuldade de codificação diferente, em função de suas características, como mostra a Tabela 4.2.

Tabela 4.2: Dificuldade de codificação das seqüências de vídeo

Seqüência	Dificuldade de codificação
CrowdRun	Difícil
ParkJoy	Difícil
DucksTakeOff	Difícil
InToTree	Fácil
OldTownCross	Fácil

Os arquivos originais requerem aproximadamente 170 GB de espaço em disco para armazenamento, e podem ser obtidos a partir da página de internet do grupo VQEG, através do roteiro da Seção B.2.1²⁰.

Uma vez copiados os arquivos originais em formato sgi16 para um disco local, é necessário convertê-los para o formato I420, o que é realizado pelo roteiro da Seção B.2.2, que utiliza a ferramenta ImageMagick (comando *convert*) em ambiente Linux.

¹⁹Disponível em: ftp://vqeg.its.blrdoc.gov/HDTV/SVT_MultiFormat/. Acesso em: ago. 2008.

²⁰Para replicar as investigações e executar os roteiros indicados deste ponto em diante, é necessário verificar a organização do ambiente, conforme descrito no início do Anexo B.



(a) CrowdRun



(b) ParkJoy



(c) DucksTakeOff



(d) InToTree



(e) OldTownCross

Figura 4.3: Sequências de vídeo para testes. Fonte: ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/

Tendo as seqüências originais completas em todas as resoluções disponíveis, o Roteiro B.2.3 gera versões truncadas das seqüências para reduzir o espaço em disco necessário para armazenamento e também o tempo das investigações, considerando principalmente etapas de codificação. Este roteiro gera duas versões de cada seqüência.

A primeira versão é composta por trechos exatos, copiados de cada seqüência no formato I420.

A segunda versão consiste em utilizar o aplicativo de subamostragem do JSVM para gerar as seqüências truncadas a partir da seqüência de resolução 2160p@50 no formato I420, que neste caso permanece igual à seqüência da primeira versão.

A diferença entre as versões está nos filtros de subamostragem utilizados para a redução de resolução das seqüências pela SVT e pelo JSVM. Uma análise técnica das diferenças entre os filtros, no entanto, não estava no escopo deste trabalho.

A Figura 4.4 resume as etapas de conversão utilizadas para gerar as seqüências truncadas para o ambiente de investigação. Pelos parâmetros de configuração do Roteiro B.1, as seqüências truncadas são compostas apenas por 100 quadros dos 500 quadros originais de cada seqüência, no caso de vídeo progressivo, ou por 50 quadros dos 250, no caso de vídeo entrelaçado. Além disso, foram descartados os 50 primeiros quadros, para que as seqüências truncadas, principalmente no caso do DucksTakeOff, contivessem as partes mais críticas das cenas, que no geral apresentam características constantes.

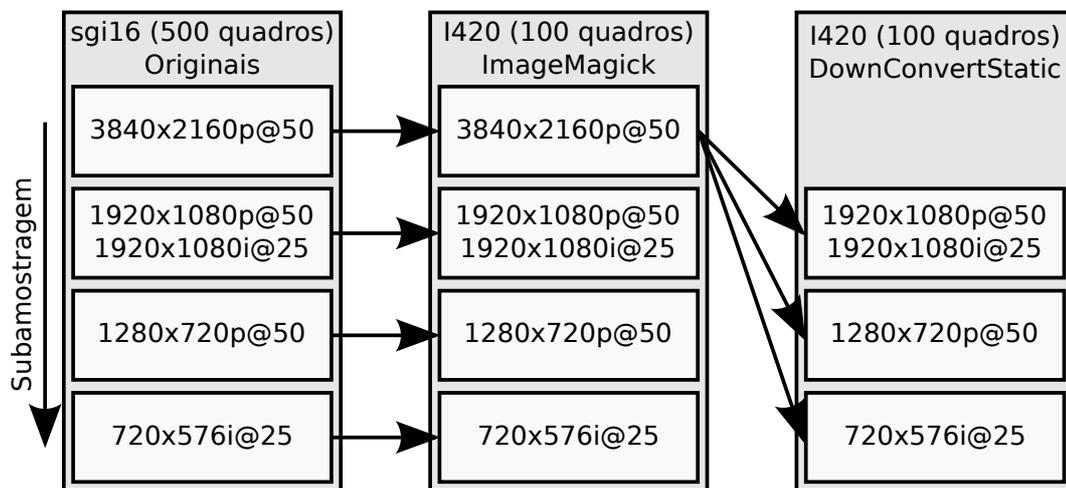


Figura 4.4: Etapas de geração das seqüências de vídeo

O último roteiro importante de preparação é o roteiro da Seção B.2.4, que serve para encapsular um conjunto de seqüências em uma camada de transporte do tipo AVI, comumente utilizada em aplicações multimídia no lugar da camada de transporte do MPEG-2. Esta etapa é necessária para a utilização do software da Ate me, que não aceita vídeo puro como entrada.

Para visualizar as seqüências no formato I420, pode ser utilizado o roteiro da Seção B.3.

4.4 Conclusão

Este capítulo apresentou considerações sobre a escalabilidade de vídeo para o SBTVD. Para isso, inicialmente, foram apresentados os projetos correlatos do JVT, do Super Hi-

Vision e da extensão de escalabilidade do sistema europeu como referência para a discussão específica da evolução de formatos de vídeo no SBTVD.

Especificamente na Seção 4.2.3, foram apresentadas as possibilidades de evolução de formato de vídeo com base na escalabilidade do SVC.

Um ambiente experimental foi proposto, considerando equipamentos para transmissão e recepção de sinal digital, codificação, decodificação, visualização e análise objetiva de vídeo.

Com base nas considerações deste capítulo, conclui-se que a aplicação do SVC pode ser viável para novos formatos no SBTVD, iniciando a evolução com o formato 1080p.

5 Levantamentos Experimentais

O objetivo deste capítulo é apresentar e discutir resultados obtidos no âmbito das considerações apresentadas no Capítulo 4, em função de experimentos realizados no ambiente de investigação implementado.

Especificamente, serão discutidas as seguintes questões: a ocupação atual do espectro da cidade de São Paulo; algumas características dos conteúdos transmitidos; algumas características das seqüências utilizadas até o momento para as investigações experimentais; as evidências indicativas de viabilidade de uso do SVC, incluindo evidências de suas premissas e a vantagem da escalabilidade sobre o método *simultcast*.

5.1 Ocupação do Espectro da Cidade de São Paulo

Primeira cidade a inaugurar as transmissões digitais de televisão terrestre no Brasil, a cidade de São Paulo, por seu contingente populacional e poder econômico, naturalmente deverá ser uma das cidades onde haverá maior demanda pelo espectro reservado para o SBTVD.

Para se ter uma estimativa do potencial de uso desse espectro, o primeiro levantamento realizado no ambiente experimental foi o de como o espectro digital estava sendo utilizado até Janeiro de 2009. Para este levantamento, foram utilizados os dois receptores digitais e o software MTS400 da Tektronix, apresentados no Capítulo 4.

Durante as investigações, a partir de Setembro de 2008, observou-se que três novos canais puderam ser recebidos, e que ao todo, em Janeiro, já havia 13 canais transmitindo sinal digital em São Paulo. Todos os canais em operação estavam configurados para transmissão hierárquica com duas camadas, uma destinada a receptores portáteis, com apenas um segmento, e outra destinada a receptores fixos, composta pelos 12 segmentos restantes. Nenhum canal, durante o período de observação, estava configurado com a terceira camada hierárquica de transmissão habilitada.

A Tabela 5.1 apresenta o resumo das informações referentes à camada hierárquica de 12 segmentos, destinada aos receptores fixos, para cada canal em operação em São Paulo.

Tabela 5.1: Configuração de transmissão para receptores fixos em São Paulo

Canal	Freq. (MHz)	IG ^a	Modulação	CC ^b	Taxa ^c (Mbps)
15	479,143	1/16	64QAM	3/4	17,84280
17	491,143	1/16	64QAM	3/4	17,84280
18	497,143	1/16	64QAM	3/4	17,84280
20	509,143	1/16	64QAM	3/4	17,84280
22	521,143	1/16	16QAM	3/4	11,89512
23	527,143	1/16	16QAM	3/4	11,89512
24	533,143	1/16	64QAM	3/4	17,84280
28	557,143	1/16	64QAM	3/4	17,84280
29	563,143	1/16	64QAM	3/4	17,84280
30	569,143	1/8	16QAM	1/2	7,48956
31	575,143	1/8	64QAM	3/4	16,85148
33	587,143	1/16	QPSK	3/4	5,94756
63	767,143	1/16	64QAM	5/6	19,82532

^aIG: Intervalo de guarda.

^bCC: Código Convolutacional.

^cTaxa máxima permitida para o TS da camada de 12 segmentos.

Alguns canais, embora identificados pelo receptor da Fujitsu, que foi utilizado para sintonia neste levantamento, não puderam ter suas informações de modulação lidas pelo aplicativo do fabricante. As condições dos sinais na Cidade Universitária, onde foram feitas as aquisições, apresentaram variações significativas em função de condições climáticas, de posicionamento da antena de recepção e, possivelmente, de variações nas condições das transmissoras. Este levantamento foi repetido algumas vezes para se obter o máximo possível de informações. Nas repetições, não foram observadas variações quanto à configuração da camada física dos canais em operação. As informações dos canais 17, 30 e 33 foram cedidas pelo Prof. Guido Stolfi por medidas realizadas na Universidade Mackenzie.

No resultado consolidado da Tabela 5.1, verifica-se que a maioria das emissoras em operação optaram pela transmissão dos 12 segmentos da programação principal com modulação 64QAM, intervalo de guarda de 1/16 e código convolutacional de 3/4. Esses parâmetros determinam uma taxa útil de 17,84280 Mbps, disponível para transmissão de áudio, vídeo e dados. Os dois casos extremos identificados neste levantamento foram o canal 33, com uma taxa útil de apenas 5,94756 Mbps, e o canal 63, com 19,82532 Mbps. Os 13 canais em operação representam uma ocupação de apenas 21% do espectro reservado para o SBTVD, ou seja, ainda há 79% do espectro disponível para novas emissoras e, possivelmente, novos serviços.

Além das características da camada física de transmissão, foram realizadas algumas aquisições sistemáticas de trechos dos fluxos de bits transmitidos pelas emissoras, considerando recepção completa dos 13 segmentos da modulação. Com essas aquisições, realizadas com o receptor da Alps e com o RTX100B, foi possível identificar algumas características dos programas alocados pelas emissoras nos fluxos transmitidos.

Dessa forma, observou-se que algumas emissoras haviam reservado, além do programa principal (considerado o programa de maior resolução ou, equivalentemente, o programa identificado como primário no TS) e do programa para receptores portáteis, pelo menos um programa adicional em resolução padrão (SDTV). Em alguns casos, este programa adicional, apesar de reservado, não apresentava conteúdo sendo transmitido. Em outros casos, o conteúdo do programa secundário era o mesmo que estava sendo transmitido no programa principal. Como em ambos os casos o programa secundário estava reservado na mesma camada de modulação da programação principal, pelas considerações apresentadas sobre os receptores do SBTVD no Capítulo 4, o programa secundário seria dispensável na transmissão, representando taxa disponível para outras aplicações e serviços. É por este motivo que a Tabela 5.2 apresenta apenas as características do programa principal, que podem ser confrontadas com as características da Tabela 5.1, para a compreensão da real utilização dos canais digitais.

Deve-se abrir aqui uma ressalva para o canal 24, no qual observou-se diferença de programação entre os canais primário e secundário, indicando a opção por multiprogramação de fato. No geral, entretanto, ficou evidente que o foco principal das emissoras tem mesmo sido a transmissão em alta resolução.

Outra ressalva quanto aos dados colhidos refere-se à utilização de formatos de vídeos não contemplados pelas especificações do SBTVD. Esta situação foi ainda observada no canal 30, com resolução 352x480 no programa principal, durante os últimos levantamentos.

De forma geral, observa-se na Tabela 5.2 que a taxa de bits utilizada para o conteúdo de vídeo do programa principal representa de 91,7 a 99,6% da taxa total do programa, com média de ocupação de 98,0%. Embora ainda não estejam sendo transmitidas aplicações interativas, por falta da normalização correspondente, este tipo de conteúdo não deverá ter um impacto médio muito significativo na atual distribuição de taxa. Considerando-se, portanto, de forma conservadora, entre 90 e 95% da taxa do programa principal e que tipicamente é obtida uma taxa de 17,84 Mpbs na camada de 12 segmentos, com a configuração atual, a maior parte das emissoras poderia utilizar entre 16 e 17 Mbps de taxa para o fluxo do vídeo principal.

Tabela 5.2: Características dos fluxos principais de vídeo das emissoras de São Paulo

Canal Freq. (MHz)	24/10/08 20h ^a	18/12/08 16h	19/12/08 8h	06/01/09 22h	07/01/09 18h
15 (479,143)	1080i29,97 ^b	1080i29,97	1080i29,97	—	1080i29,97
	14,385 ^c	14,385	14,385	—	14,385
	14,122 ^d	14,123	14,123	—	14,122
17 (491,143)	1080i29,97	1080i29,97	1080i29,97	—	1080i29,97
	8,226	10,256	10,255	—	10,256
	8,188	10,219	10,219	—	10,219
18 (497,143)	1080i29,97	1080i29,97	1080i29,97	1080i29,97	1080i29,97
	16,336	9,860	15,736	15,577	15,758
	15,515	9,040	15,518	15,360	15,538
20 (509,143)	1080i29,97	1080i29,97	1080i29,97	1080i29,97	1080i29,97
	15,909	15,909	15,910	15,911	15,910
	15,767	15,767	15,767	15,767	15,767
22 (521,143)	1080i29,97	—	—	—	1080i29,97
	9,876	—	—	—	9,803
	9,774	—	—	—	9,700
23 (527,143)	1080i29,97	1080i29,97	1080i29,97	1080i29,97	1080i29,97
	9,869	9,805	9,804	9,815	9,722
	9,735	9,670	9,669	9,681	9,587
24 (533,143)	480i29,97	480i29,97	480i29,97	1080i29,97	1080i29,97
	7,889	7,873	7,876	7,787	7,886
	7,755	7,739	7,742	7,752	7,752
28 (557,143)	1080i29,97	1080i29,97	1080i29,97	1080i29,97	1080i29,97
	12,538	12,540	12,539	12,551	12,538
	11,831	11,831	11,831	11,831	11,831
29 (563,143)	1080i29,97	1080i29,97	1080i29,97	1080i29,97	1080i29,97
	8,073	11,511	11,247	11,490	11,511
	7,808	11,245	10,982	11,224	11,245
30 (569,143)	480i29,97	480i29,97	—	480i29,97	480i29,97
	3,966	3,733	—	3,733	3,730
	3,571	3,601	—	3,601	3,598
31 (575,143)	1080i29,97	1080i29,97	1080i29,97	1080i29,97	1080i29,97
	15,595	15,594	15,593	15,595	15,594
	15,329	15,328	15,328	15,328	15,328
33 (587,143)	—	—	—	—	480i29,97
	—	—	—	—	5,308
	—	—	—	—	5,158
63 (767,143)	—	—	—	1080i29,97	1080i29,97
	—	—	—	6,870	3,404
	—	—	—	6,736	3,338

^aHorário aproximado em que foram feitas as aquisições dos fluxos de cada canal disponível e com recepção aceitável no momento da aquisição.

^bFormato do vídeo principal.

^cTaxa média, em Mbps, do programa principal incluindo áudio e vídeo.

^dTaxa média, em Mbps, apenas do fluxo de vídeo do programa principal.

Considerando-se ainda que alguns programas de alta resolução (formato 1080i@60) já estão sendo transmitidos a taxas tão baixas quanto 8 ou 9 Mpbs, e que embora sem rigor de análise, esses programas apresentem boa qualidade visual, nota-se potencial de exploração para transmissão escalável de vídeo dentro de um canal digital.

Por fim, vale ressaltar que o formato de vídeo mais utilizado para a transmissão do programa principal no SBTVD tem sido o 1080i, embora nem sempre o material transmitido tenha sido gerado nesse formato. Além disso, alguns canais apresentaram variações quanto à transmissão de formatos e taxas na programação, o que evidencia a flexibilidade de operação do sistema, que pode ser reconfigurado durante a operação. Juntando-se os fatos, a escalabilidade mostra-se eficiente para a evolução do sistema, pois a geração de conteúdo em maiores resoluções ocorrerá de forma gradual, como tem sido com o 1080i. Dessa forma, a taxa remanescente do canal, quando o material transmitido tiver uma das resoluções já suportadas, poderá ser utilizada para outros fins.

5.2 Qualidade de Codificação

Conforme apresentado no Capítulo 4, para se ter referências de validação das premissas da codificação escalável e também para acompanhamento da evolução de ferramentas de codificação, alguns levantamentos de medidas objetivas de qualidade de vídeo foram realizados no ambiente experimental. Esta seção apresenta os resultados obtidos na comparação entre os métodos de subamostragem utilizados para gerar as seqüências de vídeo a partir dos originais do SVT; apresenta resultados referentes à comparação entre estruturas de compressão tradicionalmente utilizadas com o H.264/AVC e a estrutura hierárquica por quadros do tipo B, utilizada na codificação escalável; apresenta resultados de comparação entre configurações distintas para codificação B-hierárquica; apresenta também resultados de comparação entre a codificação escalável e o método de *simulcast*.

5.2.1 Métodos de Subamostragem

Como foram utilizados dois métodos distintos de subamostragem para gerar as seqüências de vídeo no ambiente experimental, inicialmente foi feita uma comparação por inspeção visual para se determinar quais seqüências seriam utilizadas nos demais experimentos. A Figura 5.1 apresenta, como exemplo, duas amostras da mesma imagem da seqüência CrowdRun em 720p, uma gerada a partir da seqüência original do SVT com o software ImageMagick e a outra gerada pela ferramenta de subamostragem do JSVM.



(a) Filtragem de redução de tamanho em RGB e (b) Filtragem de redução aplicada diretamente na conversão entre RGB e YUV com a ferramenta seqüência YUV de maior resolução utilizando a convert do pacote ImageMagick ferramenta de conversão do JSVM

Figura 5.1: Exemplos de imagens obtidas nos processos de subamostragem para geração das seqüências de vídeo

Observou-se, de forma geral, que as imagens das seqüências geradas pelo ImageMagick apresentaram um pouco mais de nitidez. Por esse motivo, embora o manual do JSVM recomende o uso de sua própria ferramenta de subamostragem para gerar as seqüências de menor resolução para a escalabilidade espacial, escolheu-se utilizar as seqüências geradas com o processo do ImageMagick nos experimentos. As demais seqüências, por questão de tempo, poderão ser utilizadas em trabalhos futuros, para avaliação do impacto das diferenças apontadas na eficiência de codificação.

Tabela 5.3: PSNR das seqüências de teste para as resoluções 720p, 1080i, 1080p.

Seqüência	PSNR (dB)		
	720p	1080i	1080p
CrowdRun	32,12	30,49	36,61
ParkJoy	31,60	29,82	35,96
DucksTakeOff	35,39	30,56	39,33
InToTree	38,69	36,65	41,65
OldTownCross	37,29	34,59	40,48

Para se ter uma medida objetiva da diferença entre os métodos de subamostragem, a Tabela 5.3 apresenta os valores PSNR obtidos entre as seqüências para as resoluções disponíveis. Observa-se que a diferença no processo de subamostragem é mais significativa para as seqüências de menor resolução, que acumulam mais passagens de filtragem. As diferenças em 1080p, próximas de 40 dB, tendem a ser imperceptíveis.

5.2.2 Estruturas Tradicionais e Estrutura B-Hierárquica

Conforme apresentado no Capítulo 4, o software JSVM utiliza, necessariamente, o conceito de codificação hierárquica por quadros do tipo B para escalabilidade. Relembrando, neste tipo de codificação, figuras do tipo B podem ser utilizadas como referência para predição de movimento e, além disso, figuras que pertencem a uma determinada camada temporal não podem ser referenciadas por figuras de camadas inferiores.

Optou-se neste trabalho por se realizar uma comparação entre estruturas tradicionais de codificação e a codificação B-hierárquica para se ter uma referência sobre o possível impacto deste tipo de estrutura na eficiência da codificação das seqüências utilizadas.

Para isso, foi utilizado o codificador AteME, que é uma implementação comercial, possivelmente com otimizações similares para os dois tipos de compressão. Além disso, optou-se também por esta ferramenta pelo tempo de codificação que, em geral, é muito inferior ao do software de referência (JSVM), que realiza operações mais exaustivas para buscar maior eficiência de codificação. As estruturas consideradas na comparação foram do tipo IBBP, IBBBP e HBF.

A estrutura IBBP, que é uma estrutura tradicional, refere-se a uma seqüência que se inicia com a primeira figura em ordem de apresentação sendo codificada como tipo I (intra). Em seguida, é codificada a quarta figura, em ordem de apresentação, como tipo P (com predição simples), utilizando a primeira figura como referência. Posteriormente, são codificadas, em seqüência, a segunda e a terceira da ordem de apresentação, como figuras do tipo B (bi-preditiva), utilizando as figuras I e P como referência. Depois das quatro primeiras figuras codificadas, codifica-se a sétima em ordem de apresentação, como do tipo P, utilizando como referência apenas a figura I ou P precedente. A quinta e a sexta figuras são, então, codificadas usando como referências também somente as figuras I e P já codificadas. O processo vai se repetindo para as figuras seguintes, até que uma figura do tipo I seja forçada pelo codificador para configurar um ponto de onde o fluxo possa ser decodificado independentemente das figuras anteriores (RAP). O intervalo entre figuras do tipo I pode ser configurado na maioria dos codificadores, ou pode ser deixado livre, para que o codificador determine o melhor momento de inserir um ponto de início de decodificação em função de características da seqüência, como detecção de quebra de cena.

A estrutura IBBBP segue a estrutura anterior, apenas com a inclusão de uma figura do tipo B a mais entre as figuras do tipo I e P.

A estrutura HBF, por sua vez, é uma estrutura hierárquica como apresentado na Figura 2.9. Neste tipo de codificação, também é possível impor ao codificador o intervalo entre figuras do tipo I. Além disso, apesar de não fazer parte da norma H.264/AVC, convencionou-se chamar de Grupo de Figuras (*Group of Pictures*) (GOP), neste tipo de estrutura, o conjunto formado pelas figuras do tipo B, seguintes a uma figura do tipo I ou P, juntamente com a primeira figura I ou P posterior às figuras tipo B consideradas. Este conceito foi utilizado nos experimentos para definir, de forma inequívoca, a estrutura HBF em uso, juntamente com o intervalo entre figuras do tipo I.

Na prática, tanto o formato das seqüências tradicionais quanto o tamanho do GOP na codificação HBF podem variar durante a codificação em função das características locais das seqüências de vídeo. Neste trabalho, porém, como as seqüências não apresentam cortes de cena, com preservação das características da filmagem durante toda a seqüência, foram utilizadas sempre estruturas fixas de codificação.

Para a avaliação comparativa entre as estruturas citadas foram utilizados os arquivos descritos na Seção B.4. Nestes arquivos, o intervalo entre figuras do tipo I foi configurado para 25 em todas as estruturas e foi ativado o controle de taxa para 4, 8, 12 e 16 Mbps. A estrutura HBF foi configurada com GOP de tamanho 4, para preservar proporção de figuras I, P e B compatível com as outras estruturas. Além disso, foram codificadas e analisadas quanto ao valor médio de PSNR as seqüências no formato 720p e 1080p, que embora tenham sido produzidas com taxa de 50 quadros por segundo, foram codificadas na taxa de 59,94 quadros por segundo, sem interpolação de quadros. Não foi possível realizar o mesmo procedimento com as seqüências em 1080i, por dificuldade em gerar os fluxos de entrada encapsulados em AVI com entrelaçamento ativo através do MEncoder e também pela impossibilidade de forçar a ferramenta de codificação da Ate a tratar a seqüência de entrada como vídeo entrelaçamento. Este formato de vídeo será tratado em trabalhos futuros.

A Tabela 5.4 apresenta os resultados obtidos para a codificação das seqüências nos formatos 720p e 1080p. Observa-se que as seqüências consideradas fáceis (InToTree e OldTownCross) apresentam qualidade média muito superior (em torno de 8 dB) à qualidade média das seqüências consideradas difíceis, independentemente do tipo de estrutura, para taxas equivalentes. Quanto às estruturas, observa-se que a codificação hierárquica apresenta um pequeno ganho de qualidade praticamente em todas as condições.

Além dos valores médios da tabela, a Figura 5.2 apresenta os gráficos de qualidade quadro a quadro das seqüências 720p, com taxa de 12 Mbps para as seqüências difíceis

Tabela 5.4: Qualidade PSNR média das seqüências 720p/1080p com o codificador Ateme

Seqüência	Taxa (kbps)	PSNR(dB)					
		720p			1080p		
		HBF	IBBP	IBBBP	HBF	IBBP	IBBBP
CrowdRun	4000	23,57	23,64	23,40	22,37	22,36	22,27
	8000	26,36	26,29	25,94	25,02	25,05	24,70
	12000	27,90	27,79	27,57	26,71	26,69	26,37
	16000	29,18	29,03	28,74	27,93	27,80	27,52
ParkJoy	4000	22,80	22,68	22,51	21,63	21,44	22,27
	8000	25,01	24,85	24,63	23,60	23,39	23,17
	12000	26,54	26,34	26,10	24,98	24,72	24,53
	16000	27,66	27,50	27,26	25,99	25,74	25,46
DucksTakeOff	4000	24,22	24,07	23,85	23,26	23,06	23,02
	8000	26,70	26,45	26,24	24,78	24,66	24,47
	12000	28,27	27,96	27,71	26,46	26,33	26,08
	16000	29,24	28,98	28,74	27,43	27,21	26,97
InToTree	4000	33,91	33,90	33,92	31,90	32,01	31,97
	8000	35,38	35,31	35,31	33,34	33,33	33,32
	12000	36,12	35,97	36,05	33,90	33,86	33,87
	16000	36,71	36,55	36,64	34,31	34,23	34,24
OldTownCross	4000	34,19	34,11	34,16	31,62	31,65	31,50
	8000	35,72	35,66	35,64	33,20	33,10	33,12
	12000	36,45	36,40	36,38	33,67	33,58	33,60
	16000	36,94	36,89	36,86	34,00	33,90	33,92

e 8 Mbps para as seqüências fáceis. A Figura 5.3, por sua vez, apresenta os gráficos de qualidade quadro a quadro das seqüências 1080p, com taxa de 16 Mbps para as seqüências difíceis e 8 Mbps para as seqüências fáceis.

Observa-se que o perfil de variação de qualidade de codificação por tipo de imagem é bastante similar entre as diferentes estruturas avaliadas, pois as figuras do tipo I apresentam qualidade expressivamente superior às demais (picos de qualidade), o que resulta em uma qualidade objetiva média similar entre as estruturas, como visto nas tabelas.

5.2.3 Estruturas B-Hierárquicas

Uma vez definida a estrutura B-hierárquica como base para a codificação escalável, existem ainda diversos parâmetros que podem ser ajustados para a codificação das seqüências de vídeo. Entre os parâmetros, a distância entre figuras do tipo I e o tamanho do GOP são especialmente importantes para a codificação visando a radiodifusão. O intervalo entre figuras do tipo I relaciona-se diretamente ao tempo necessário para sintonizar um

canal, pois o receptor só é capaz de iniciar a decodificação do fluxo de vídeo a partir destas figuras, e ainda com a condição de que figuras anteriores não sejam mais utilizadas como referência. Além disso, o tamanho utilizado para o GOP influencia diretamente na quantidade mínima de figuras necessárias como referência na decodificação, e, portanto, determina indiretamente o nível mínimo que o decodificador deve suportar.

Para a radiodifusão no SBTVD, foi recomendado que as emissoras procurem manter intervalo de meio segundo entre pontos de início de decodificação para não causar desconforto ao usuário. Até um segundo de intervalo ainda é tolerável, embora já signifique uma diferença perceptível.

Quanto aos níveis de decodificação, conforme apresentado no Capítulo 3, o SBTVD atualmente exige suporte ao nível 4.0. Além disso, conforme discutido no Capítulo 4, o nível mais adequado para o primeiro passo de evolução do sistema seria 4.2, que limita em 4 a quantidade de figuras de referência para codificação no formato 1080p.

Com essas considerações, optou-se pela avaliação do comportamento de codificação em camada simples das seqüências nos formatos 720p@50 e 1080p@50 com o software JSVM. Nesta avaliação os parâmetros estruturais foram variados, gerando-se combinações entre intervalos de 25 ou 50 quadros entre figuras do tipo I e tamanhos de GOP de 4 ou 8. Como as seqüências foram codificadas em suas taxas de quadros de produção, os intervalos selecionados representam 0,5 s e 1,0 s entre imagens do tipo I respectivamente. O tamanho de GOP de 8 figuras representa o máximo valor suportado pelo nível 4.2, por exigir 4 figuras de referência na codificação.

Definidos os parâmetros estruturais, optou-se pela codificação com taxa variável, pois verificou-se que o JSVM apresenta controle de taxa apenas para codificação de seqüências sem quadros do tipo B. Na codificação com taxa variável, é possível definir o coeficiente de quantização de cada camada temporal, que permanece como uma referência constante para todas as imagens da mesma camada. O coeficiente de quantização para cada tipo de macrobloco é derivado do coeficiente de referência da imagem, de forma que a codificação ocorre buscando manter a qualidade constante dentro da seqüência. Especificamente no JSVM é possível definir um coeficiente global de referência, a partir do qual o software deriva automaticamente os coeficientes para cada camada temporal, baseado em levantamentos estatísticos realizados pelas equipes que o desenvolveram. O coeficiente global de quantização foi variado de 24 a 40, com passos de 4, para as seqüências 720p e de 28 a 40, também com passos de 4, para as seqüências 1080p. A Tabela 5.5 resume os coeficientes derivados pelo JSVM para cada camada, independentemente da resolução da seqüência.

Tabela 5.5: Coeficientes derivados pelo JSVM para as camadas temporais

Coeficiente de Referência	Estrutura (GOP)	Camada temporal			
		0	1	2	3
24	4	21	25	26	—
	8	20	23	25	26
28	4	25	29	30	—
	8	24	27	29	30
32	4	29	33	34	—
	8	28	31	33	34
36	4	33	37	38	—
	8	32	35	37	38
40	4	37	41	42	—
	8	36	39	41	42

Inicialmente os roteiros de codificação que automatizam este experimento, disponíveis na Seção B.5, haviam sido configurados para permitir uma região de busca de 256 por 256 pixels durante a predição de movimento. Com essa configuração, no entanto, verificou-se que era necessário 1 GB de memória volátil para codificação das seqüências 1080p, além de cerca de 3,5 horas para a codificação de apenas 49 quadros em um computador com processador Intel Core 2 DuoTM de 2.4 GHz. Com a diminuição da região de busca na predição de movimento para uma janela de 32 por 32 pixels, a memória necessária foi reduzida para cerca de 650 MB e o tempo de codificação de 49 quadros foi reduzido para cerca de 40 minutos. O impacto desta limitação no resultado da eficiência de codificação foi desprezível, afetando apenas a partir da segunda casa decimal dos valores PSNR medidos.

Nos roteiros, a codificação das seqüências foi limitada a apenas 49 quadros, para reduzir também o tempo dos experimentos, visto que as características das seqüências mantêm-se constante.

A Tabela 5.6 resume o nível reservado pelo codificador para cada tipo de estrutura (varia em função da resolução e do tamanho de GOP somente), conforme análise dos relatórios gerados pela ferramenta de codificação.

Tabela 5.6: Níveis de codificação alocados pelo JSVM

Formato	GOP	Nível
720p	4	3.2
	8	4.0
1080p	4	4.2
	8	5.0

Observa-se que com GOP de tamanho 8 o JSVM aloca para o fluxo codificado um

nível superior ao realmente necessário. Analisando o software, verificou-se que isso acontece pelo superdimensionamento do buffer reservado para armazenamento de quadros de referência. No caso específico do formato 1080p com GOP de 8, por exemplo, o buffer reservado permite o armazenamento de 7 figuras de referência, quando seriam necessárias apenas 4, como visto anteriormente, o que permitiria a codificação no nível 4.2. Essa observação é importante do ponto de vista de requisitos para a decodificação, pois como discutido, a próxima família de receptores deve suportar somente até o nível 4.2. Não é possível determinar, sem alterar o software para realizar novamente os experimentos, se o superdimensionamento do nível tem impacto na eficiência da codificação, aumentando a qualidade para uma mesma taxa. Espera-se, no entanto, que este impacto seja mínimo, uma vez que durante a predição de movimento a maior probabilidade de referência de quadros seria dos mesmos quadros que estariam disponíveis na codificação com o nível correto.

Os resultados gráficos obtidos a partir dos levantamentos com os roteiros da Seção B.5 são apresentados da Figura 5.4 à Figura 5.9. A escala de taxa utilizada nos gráficos está baseada na relação entre alocação de bits para o fluxo codificado e a taxa de quadros das seqüências, representando uma taxa média de bits para seqüência. Cada ponto do gráfico representa um dos coeficientes de quantização utilizados como referência, sendo que o menor coeficiente resulta no ponto de maior taxa da seqüência e o maior coeficiente no ponto de menor taxa, com comportamento monotônico para os valores intermediários.

De forma geral, observa-se nestas figuras que o impacto de eficiência de codificação do intervalo entre figuras I nas medidas de qualidade PSNR e SSIM é inferior ao impacto dos demais parâmetros (tamanho do GOP e valor do coeficiente de quantização). De fato, avaliando a relação entre taxa e qualidade para os coeficientes utilizados, como exemplifica a Tabela 5.7 para a seqüência CrowdRun, verifica-se que GOP de tamanho 8 é vantajoso em relação a GOP de tamanho 4, pois apresenta relação de taxa de bits por dB de qualidade inferior no intervalo de coeficientes avaliados.

Nos gráficos, as medidas PSNR indicam ganho de qualidade com derivada ainda expressiva mesmo para taxas elevadíssimas, quando se utiliza baixos coeficientes de referência (entre 24 e 28). Já as curvas SSIM aproximam-se mais rapidamente de uma saturação, evidenciando a possível obtenção de boa qualidade subjetiva na codificação das seqüências em taxas mais factíveis em termos de implementação e compatibilidade sistêmica. As seqüências consideradas difíceis, no formato 1080p, devem apresentar ainda boa qualidade subjetiva para taxas inferiores a 20 Mbps, enquanto as consideradas fáceis podem ser satisfatórias com menos de 10 Mbps.

Tabela 5.7: Resultados da codificação da seqüência CrowdRun em camada simples com o JSVM

Formato	Qp	720p			1080p		
		Taxa (kbps)	PSNR (dB)	Taxa/PSNR (kbps/dB)	Taxa (kbps)	PSNR (dB)	Taxa/PSNR (kbps/dB)
I48 GOP4	40	6359,9347	27,47	231,542	10929,7306	28,64	381,682
	36	10696,6857	29,93	357,338	18193,7061	30,96	587,717
	32	17392,9224	32,42	536,438	29786,9878	33,21	896,893
	28	28406,9633	35,10	809,237	50440,2041	35,46	1422,27
	24	45197,1755	37,86	1193,94	—	—	—
I48 GOP8	40	5956,0327	27,43	217,101	10346,3429	28,64	361,29
	36	10126,4163	29,80	339,827	17326,4000	30,86	561,408
	32	16763,6000	32,31	518,761	28771,7061	33,13	868,417
	28	27181,7388	34,92	778,325	48003,6571	35,32	1359,07
	24	43861,9184	37,68	1163,98	—	—	—
I24 GOP4	40	6586,5143	27,52	239,349	11306,4082	28,69	394,145
	36	11022,7020	29,97	367,821	18750,7429	30,99	605,009
	32	17830,6857	32,45	549,429	30547,6408	33,23	919,412
	28	28986,6449	35,12	825,384	51521,9184	35,47	1452,53
	24	45935,4531	37,87	1213,1	—	—	—
I24 GOP8	40	6121,3224	27,46	222,936	10627,2408	28,66	370,743
	36	10363,6490	29,82	347,588	17741,7878	30,88	574,551
	32	17103,4449	32,33	529,084	29377,8612	33,14	886,576
	28	27629,5184	34,93	790,913	48778,0327	35,32	1380,84
	24	44366,8245	37,69	1177,27	—	—	—

Um resultado inesperado, que deve ser avaliado junto à implementação da ferramenta de medição SSIM utilizada (por falta de documentação), é a saturação em valores abaixo de 0.9 para as seqüências consideradas fáceis no formato 1080p. Esta baixa média de valores foi determinada por flutuações muito expressivas (em torno de 0,1) entre os valores SSIM de imagens codificadas como tipo I ou P e imagens codificadas como tipo B nestas seqüências. A ferramenta de análise SSIM deverá ser investigada em trabalhos futuros para esclarecer este comportamento.

5.3 Comparação Entre Escalabilidade e *Simulcast*

Avaliado o comportamento da codificação para as camadas individuais das seqüências de vídeo, esta seção apresenta a investigação realizada sobre a codificação escalável entre os formatos 720p e 1080p para comparação de eficiência com uma transmissão simultânea (*simulcast*) equivalente.

Para isso, em função dos resultados anteriores determinou-se que a estrutura utilizada nos testes seria configurada com intervalo de 24 entre figuras I e GOP de tamanho 8. A

escolha dos valores dos coeficientes de quantização para a codificação escalável, por outro lado, seria melhor determinada por avaliações subjetivas sobre os fluxos simples gerados, o que não estava no escopo deste trabalho. Além disso, os tempos envolvidos nos testes de codificação, mesmo com parâmetros simplificados e seqüências encurtadas, torna o processo de avaliação exaustiva bastante demorado. Por esses motivos, para as avaliações iniciais deste trabalho, optou-se pela escolha dos coeficientes de referência em função das taxas obtidas nos gráficos apresentados na Figura 5.7 para a camada de enriquecimento em 1080p e na Figura 5.6 para a camada base em 720p.

Tomando-se por referência uma taxa próxima de 16 Mbps para as seqüências difíceis em 1080p e 8 Mbps para as seqüências fáceis na mesma resolução, verifica-se, por aproximação, na Figura 5.7, que os coeficientes necessários para se obter cada seqüência codificada nas taxas citadas e os respectivos valores de PSNR seriam os apresentados na Tabela 5.8. Nesta tabela, são apresentados também os valores de coeficientes obtidos na Figura 5.6 para que as seqüências em 720p fossem codificadas como camada base, apresentando qualidade PSNR similar à da camada de enriquecimento.

Tabela 5.8: Referências das seqüências em 1080p para a codificação escalável

Seqüência	Qp (1080p)	PSNR (aprox.) (dB)	Qp (720p)
CrowdRun	38	29,8	36
Parkjoy	39	27,6	37
DucksTakeOff	36	30,7	36
InToTree	32	34,4	33
OldTown	32	34,4	34

Os roteiros utilizados para gerar as codificações escaláveis a partir dos coeficientes da Tabela 5.8 encontram-se na Seção B.6. Além das codificações escaláveis, foram gerados também os fluxos simples no formato 1080p para levantamento da taxa exata associada ao coeficiente utilizado. A Tabela 5.9 apresenta os resultados relativos à camada base da codificação escalável em 720p e ao fluxo simples em 1080p, para simular uma transmissão em *simulcast*, representada pela soma das taxas dos dois fluxos.

A Tabela 5.10 apresenta os resultados obtidos com o fluxo escalável completo, com camada base em 720p e camada de enriquecimento em 1080p. A economia de taxa da escalabilidade em relação à transmissão *simulcast*, a sobrecarga de taxa da codificação escalável em relação a uma transmissão simples em 1080p e a perda de qualidade do fluxo escalado com relação ao fluxo 1080p também são apresentados na tabela.

Verifica-se que para os parâmetros escolhidos nos testes, as seqüências difíceis apre-

Tabela 5.9: Resultados do *simulcast* de 720p e 1080p

Seqüência	Taxa 720p (kbps)	PSNR 720p (dB)	Taxa 1080p (kbps)	PSNR 1080p (dB)	<i>Simulcast</i> (kbps)
CrowdRun	10404,6531	29,68	13838,9878	29,63	24243,6409
Parkjoy	10171,2408	27,94	14475,3388	27,59	24646,5796
DucksTakeOff	8702,5306	30,09	15813,3959	30,43	24515,9265
InToTree	2625,8449	34,49	7525,4612	34,27	10151,3061
OldTown	2492,1959	34,65	6756,8163	34,36	9249,0122

Tabela 5.10: Resultados da escalabilidade de 720p para 1080p

Seqüência	SVC (720p-1080p) (kbps)	PSNR (dB)	Economia (%)	Sobrecarga (%)	Perda (dB)
CrowdRun	17339,0694	29,60	28,5	25,3	0,03
Parkjoy	17320,8898	27,55	29,7	19,7	0,04
DucksTakeOff	16921,8612	30,36	31,0	7,0	0,07
InToTree	8461,9592	34,27	16,6	12,4	0,00
OldTown	7595,8776	34,31	17,9	12,4	0,05

sentam em torno de 30% de economia de taxa em relação a uma transmissão *simulcast*, com uma perda desprezível de qualidade na medida PSNR. A economia nas seqüências fáceis é mais discreta, em torno de 17%, mas ainda assim, a escalabilidade mostra-se, em todos os casos, como solução vantajosa para evolução de um sistema onde se deseja manter o legado. Ganhos mais expressivos devem poder ser obtidos com as seqüências de menor resolução geradas pela ferramenta do próprio JSVM. Essa comparação será realizada em trabalhos futuros.

5.4 Transmissão para Receptor Legado

A análise das características da codificação escalável e suas próprias premissas de especificação evidenciam que, respeitados os parâmetros do sistema legado para a codificação da camada base, a evolução do sistema garante a compatibilidade com o sistema legado.

Para realizar testes de compatibilidade da escalabilidade no âmbito do SBTVD, foram iniciados testes com o multiplexador da Tektronix. No entanto, foram encontradas incompatibilidades entre o fluxo de vídeo codificado gerado pelo JSVM e a ferramenta de multiplexação. Embora o manual do equipamento não imponha restrições ao fluxo de vídeo codificado, respeitando-se a norma H.264/AVC, a ferramenta de multiplexação é encerrada quando se tenta incluir um fluxo de vídeo codificado pelo JSVM, com ou mesmo sem escalabilidade.

Desta forma, não foi possível comprovar experimentalmente a possibilidade de transmissão de um fluxo escalável para um receptor do SBTVD. Será necessário investigar, como trabalho futuro, junto ao fabricante do equipamento, o motivo da incompatibilidade.

5.5 Conclusão

A partir dos trabalhos iniciais realizados sobre o ambiente experimental especificado e montado, verificou-se que um ano após o início das transmissões, a ocupação do espectro digital na cidade de São Paulo apresenta disponibilidade expressiva para novos formatos de vídeo e/ou novos serviços.

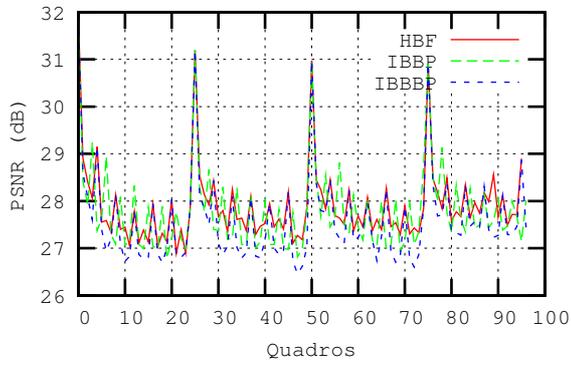
Duas frentes de trabalho são derivadas a partir das evidências sobre a ocupação do espectro: a primeira é a possibilidade de avanços técnicos e a natural otimização dos parâmetros de codificação do H.264/AVC. Neste caso, é possível afirmar que taxas maiores de compressão e conseqüentemente maior sobra de espectro serão observados num futuro próximo. A segunda frente de trabalho é de otimização da camada física de transmissão. Muitos amplificadores foram instalados e configurados de forma empírica pelas emissoras, possibilitando investigação nesta área, já com foco adicional na escalabilidade.

Neste capítulo também foi realizada uma comprovação experimental de resultados de codificação, partindo da eficiência da estrutura B-hierárquica. Além disso, foram avaliadas possíveis configurações desta estrutura para a comparação entre a escalabilidade SVC da norma H.264/AVC e o *simulcast*.

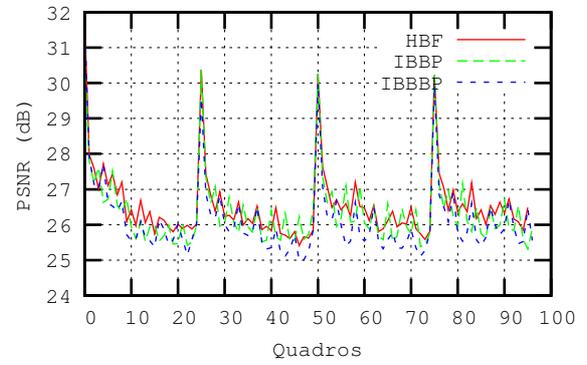
Apesar de resultados similares terem sido apresentados na literatura, como citado no Capítulo 2 e no Capítulo 4, esta validação experimental é importante para o domínio do exercício destes testes para futuros testes mais exaustivos. Os roteiros utilizados no ambiente experimental estão disponíveis no Anexo B desta dissertação.

A principal conclusão destes testes é que a escalabilidade apresenta vantagem significativa sobre o *simulcast*, o que evidencia objetivamente a possibilidade de adoção futura pelo SBTVD de formatos superiores ao 1080i. Entretanto deve-se considerar fortemente trabalhos futuros complementares a esta dissertação na área de testes subjetivos de vídeo para que as evidências aqui apresentadas sejam validadas sobre uma perspectiva mais ampla e rigorosa.

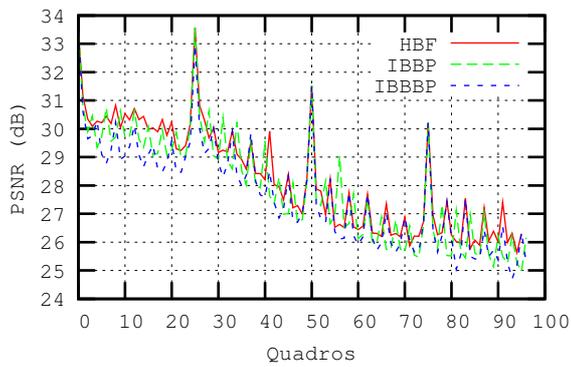
Outra frente futura de trabalhos é aprimorar os testes sobre o software de referência num esforço de superação de suas limitações, como as apontadas neste capítulo.



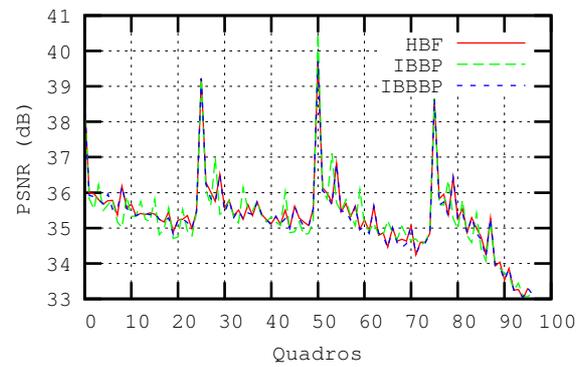
(a) CrowdRun 720p 12Mbps



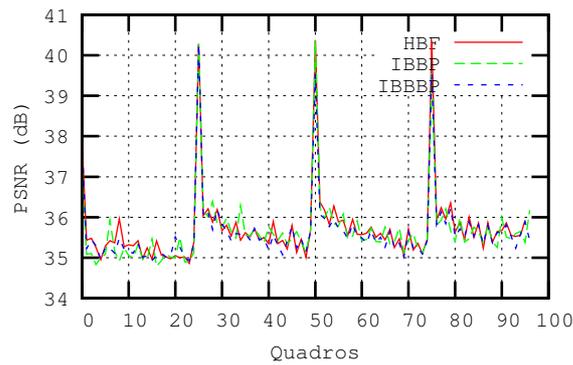
(b) ParkJoy 720p 12Mbps



(c) DucksTakeOff 720p 12Mbps

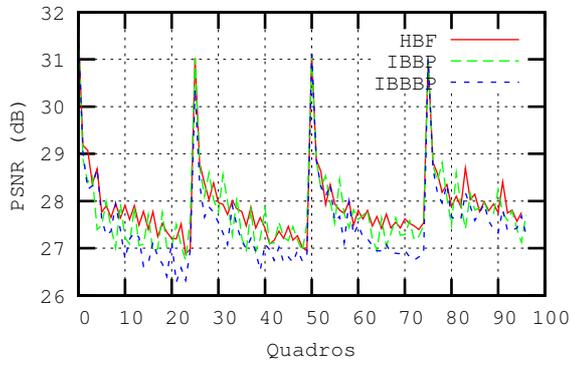


(d) InToTree 720p 8Mbps

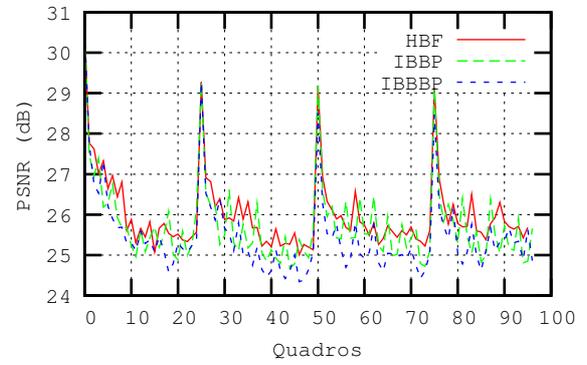


(e) OldTownCross 720p 8Mbps

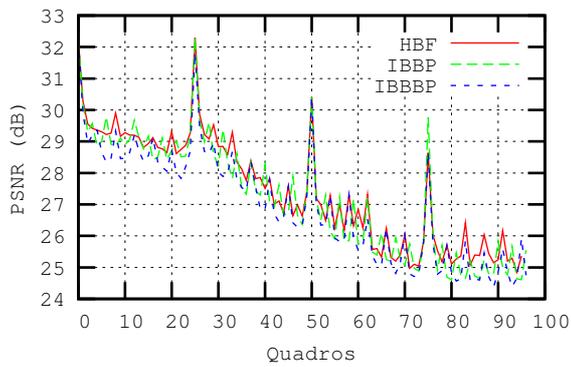
Figura 5.2: Comparação de qualidade entre diferentes estruturas de codificação (Formato 720p)



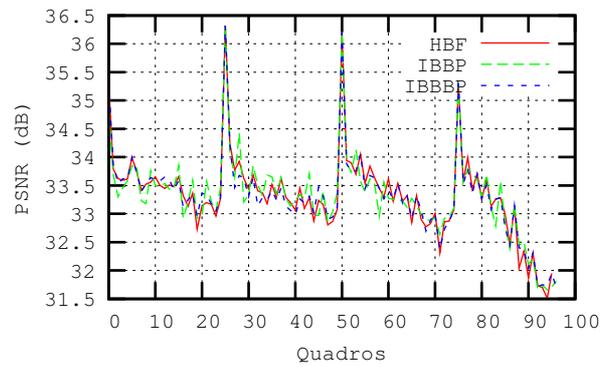
(a) CrowdRun 1080p 16Mbps



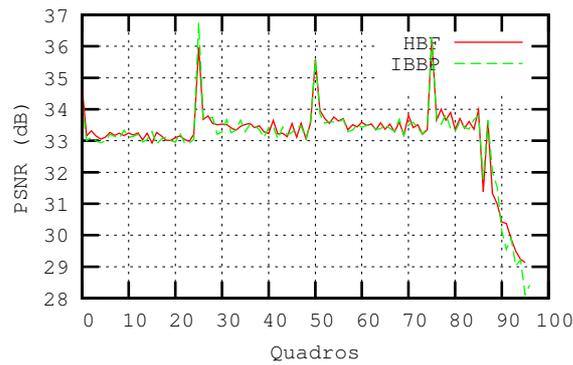
(b) ParkJoy 1080p 16Mbps



(c) DucksTakeOff 1080p 16Mbps

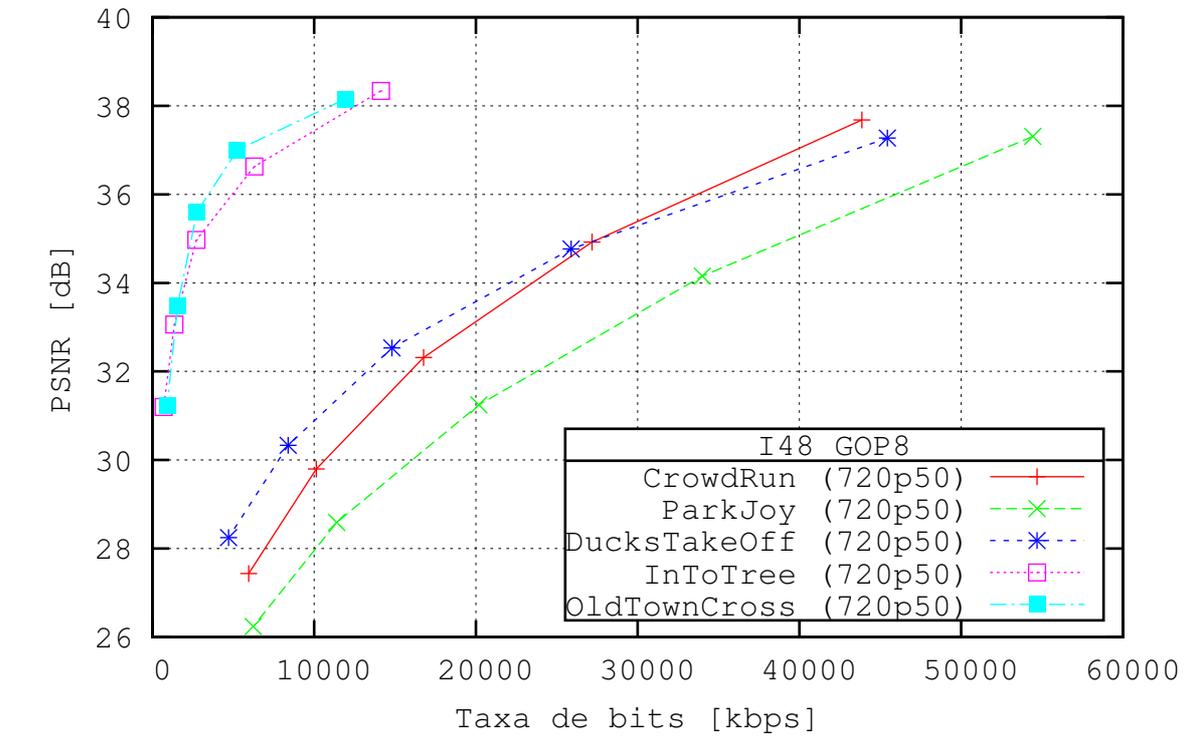


(d) InToTree 1080p 16Mbps

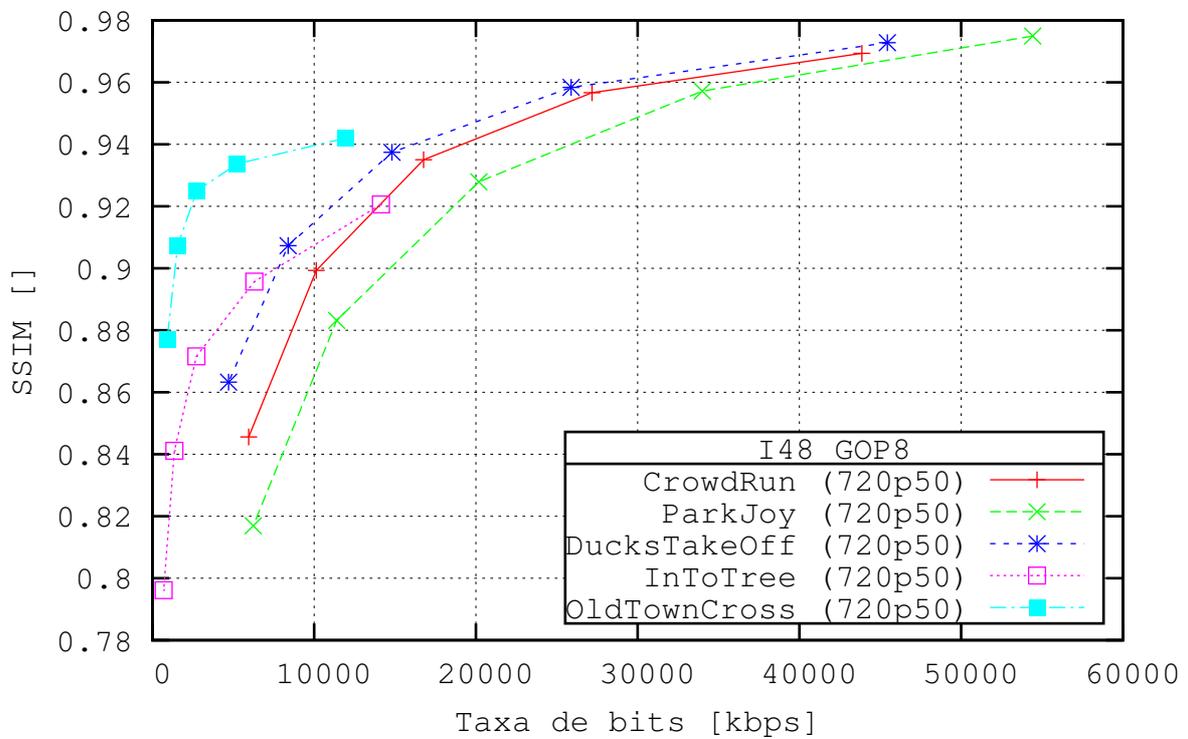


(e) OldTownCross 1080p 8Mbps

Figura 5.3: Comparação de qualidade entre diferentes estruturas de codificação (Formato 1080p)

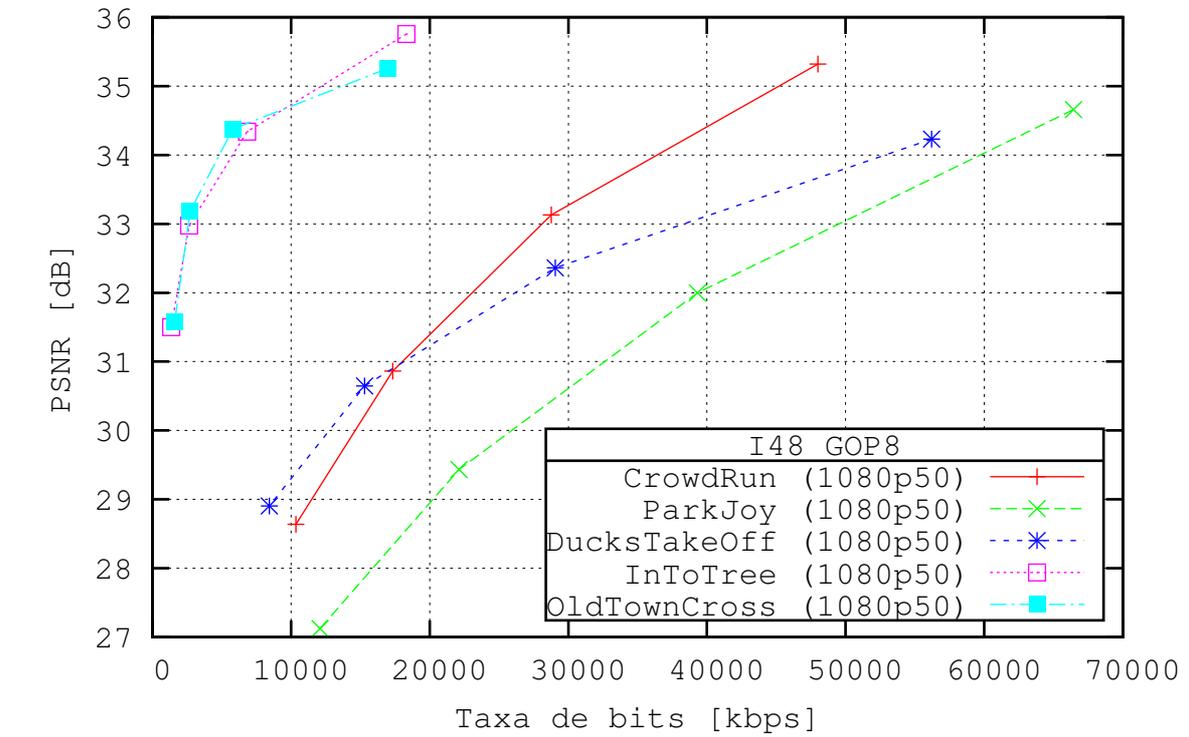


(a) PSNR

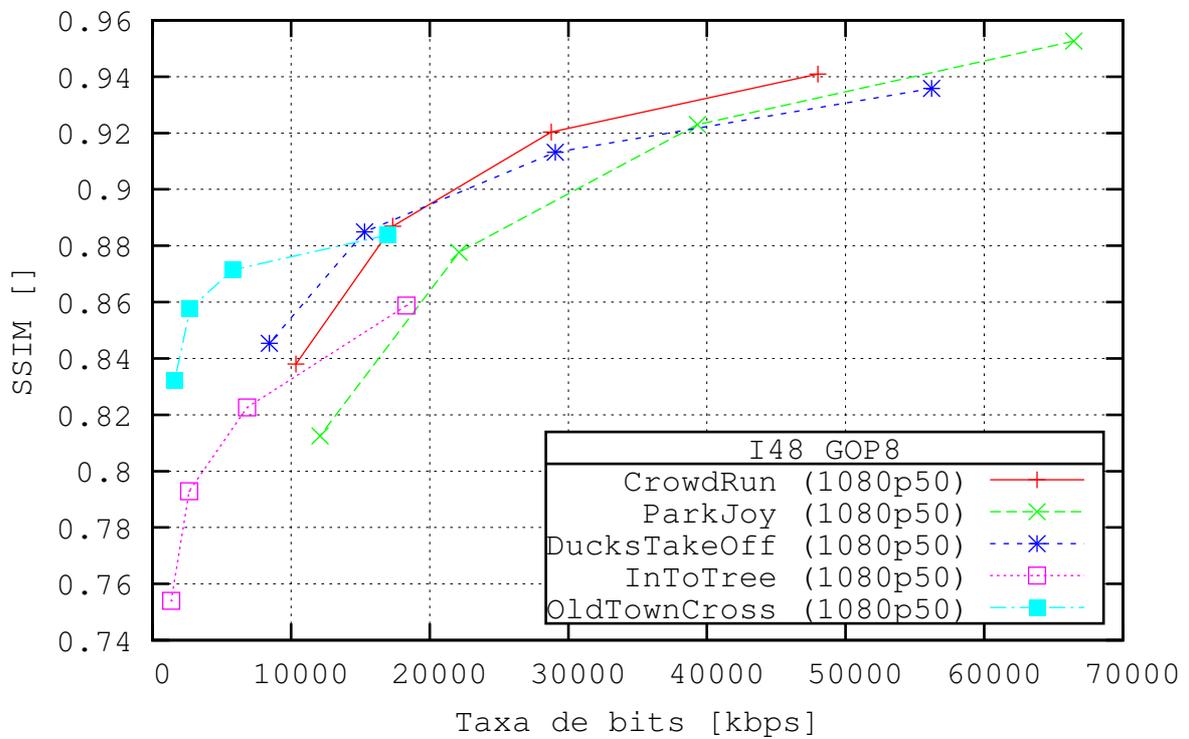


(b) SSIM

Figura 5.4: Resultados da codificação por camada simples (720p, I=48, GOP=8)

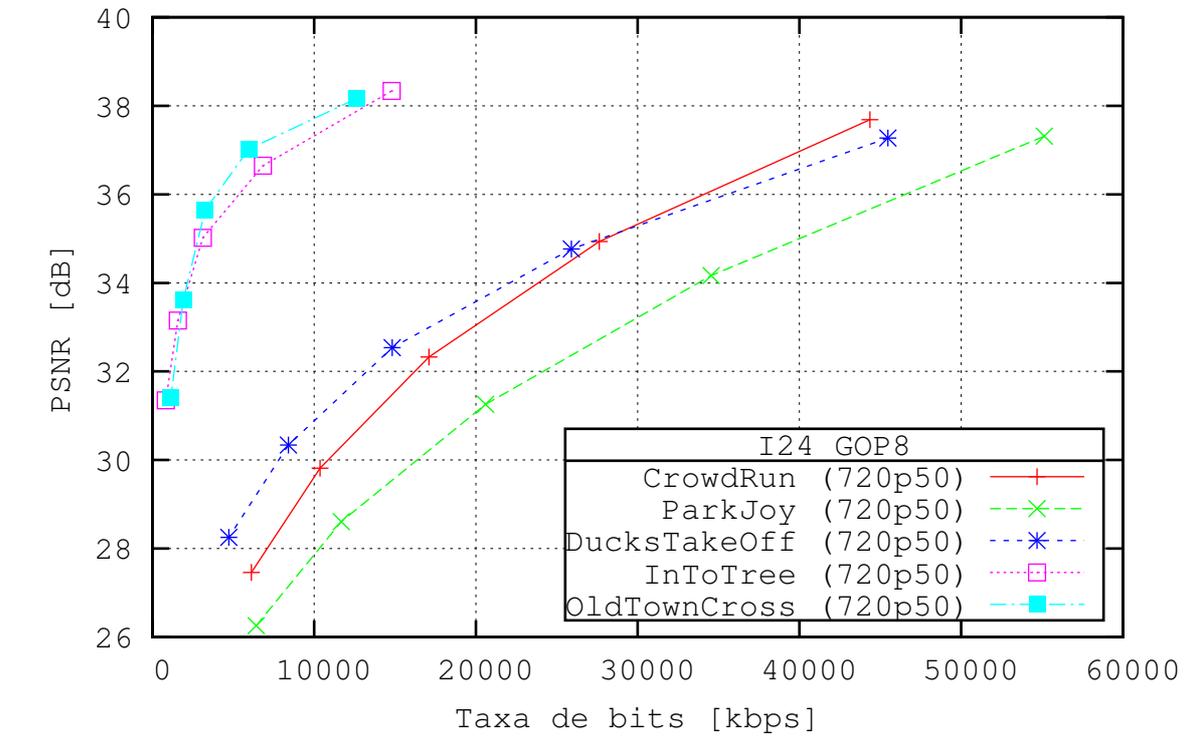


(a) PSNR

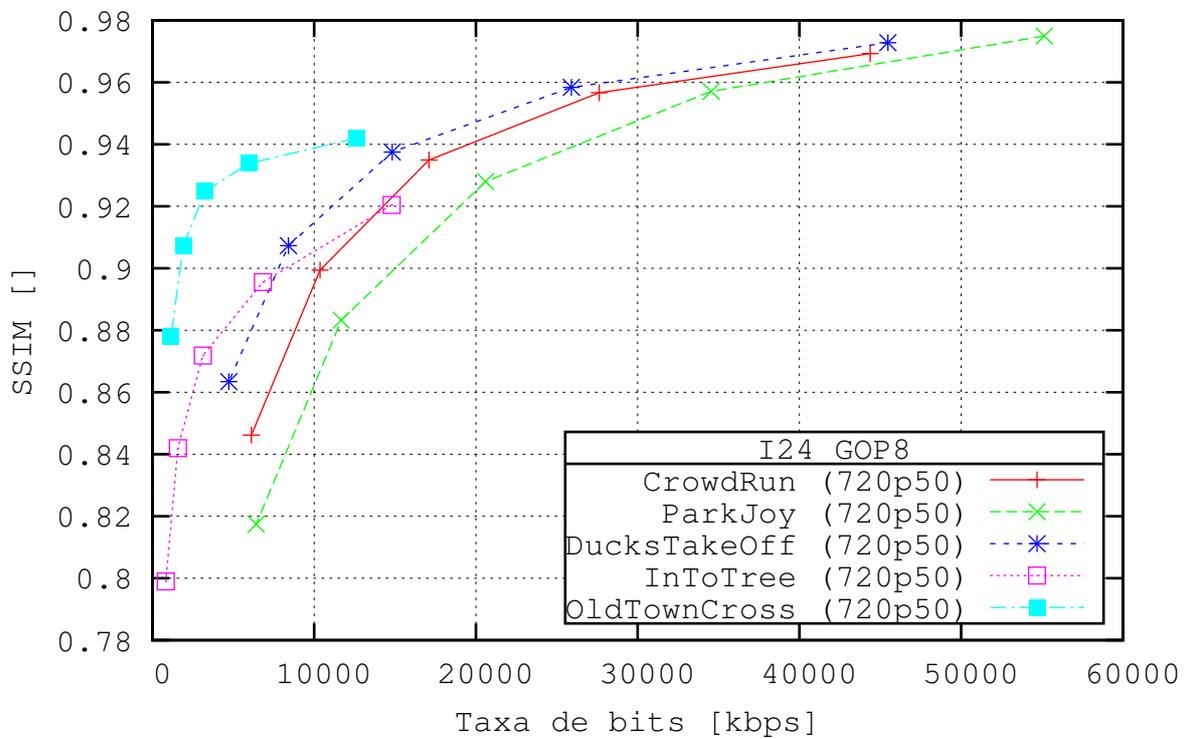


(b) SSIM

Figura 5.5: Resultados da codificação por camada simples (1080p, I=48, GOP=8)

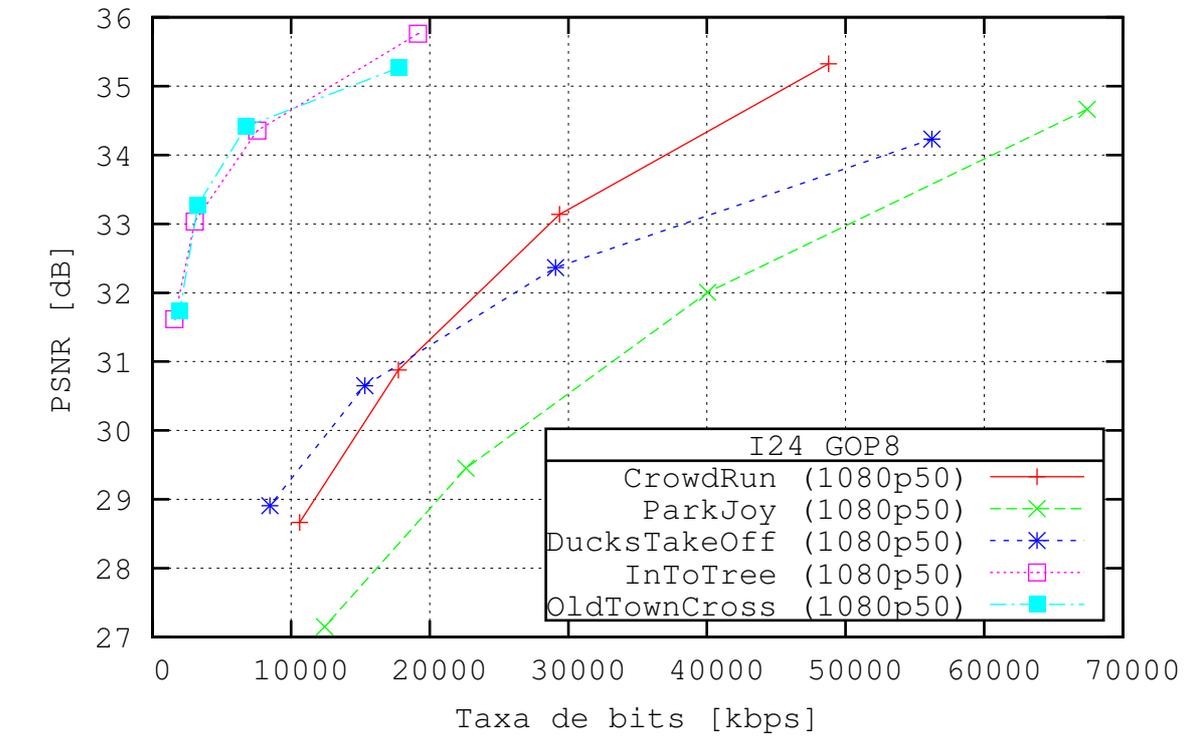


(a) PSNR

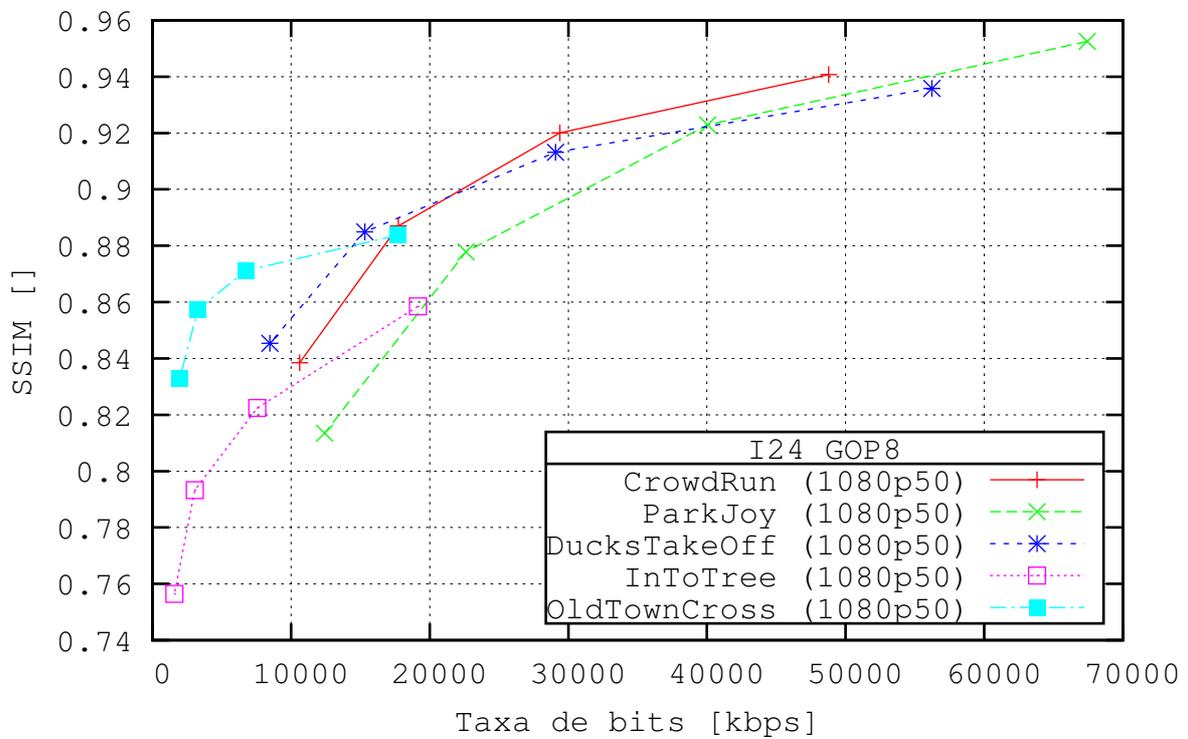


(b) SSIM

Figura 5.6: Resultados da codificação por camada simples (720p, I=24, GOP=8)

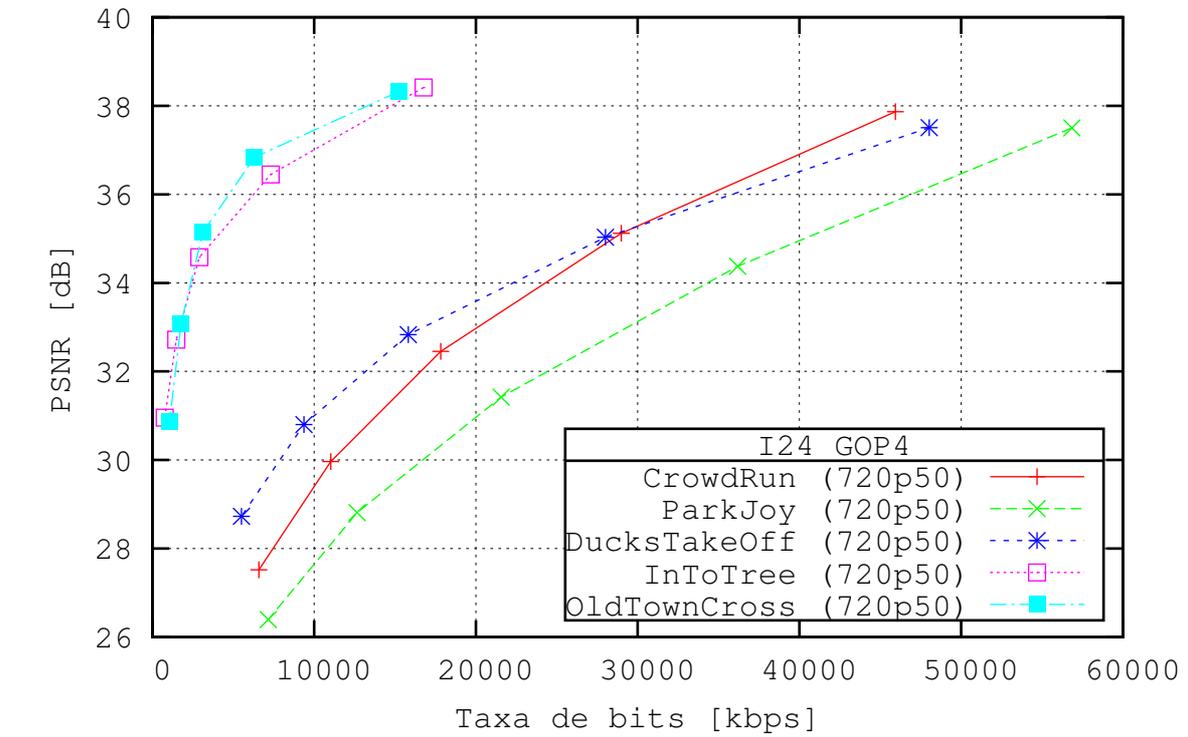


(a) PSNR

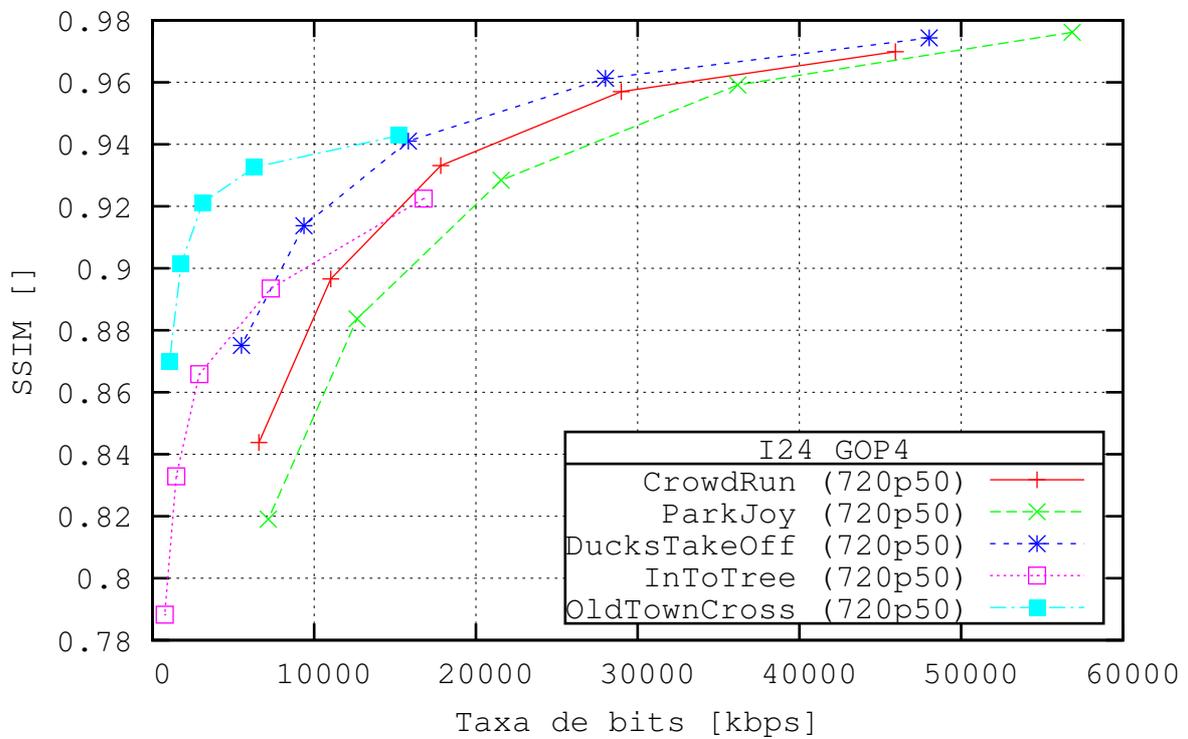


(b) SSIM

Figura 5.7: Resultados da codificação por camada simples (1080p, I=24, GOP=8)

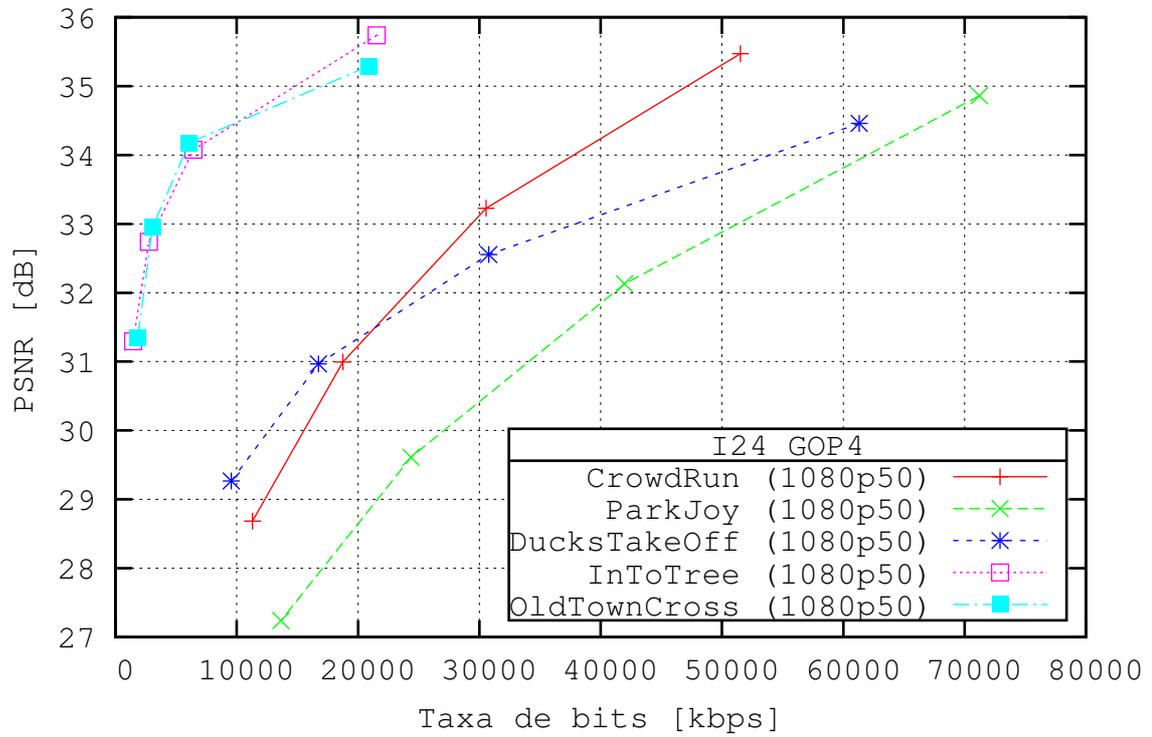


(a) PSNR

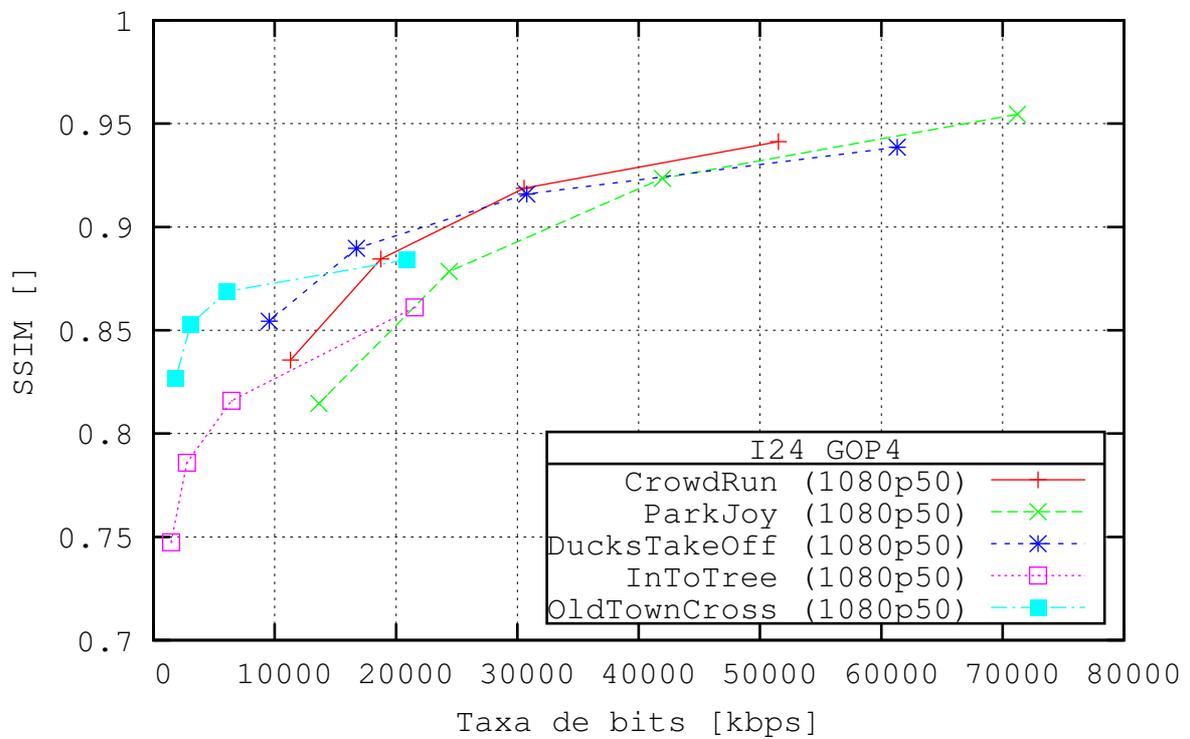


(b) SSIM

Figura 5.8: Resultados da codificação por camada simples (720p, I=24, GOP=4)



(a) PSNR



(b) SSIM

Figura 5.9: Resultados da codificação por camada simples (1080p, I=24, GOP=4)

6 Conclusões

Os primeiros estudos para a introdução da Televisão Digital no Brasil remetem ao início da década de 90, quando considerava-se a adoção do padrão MPEG-2 para a camada de transporte do sistema e também para a codificação de áudio e vídeo. No entanto, desde de 2004, quando as ações decisivas para o modelo brasileiro tiveram início com o apoio do governo, o grupo de pesquisa do LSI defendeu fortemente a adoção do padrão H.264/AVC, inclusive com suporte ao formato 1080p@60, para a codificação de vídeo no sistema. Conforme discutido no Capítulo 4, algumas limitações tecnológicas momentâneas impediram a adoção deste formato. Neste processo, o autor participou das equipes técnicas de pesquisa, discussão e especificação do sistema, especialmente na área de codificação de vídeo.

A experiência principal do autor, inicialmente, estava na arquitetura de receptores digitais, com ênfase no fluxo de decodificação multimídia, que é um elemento essencial da arquitetura, particularmente apresentando alta exigência de requisitos computacionais de desempenho. Com os trabalhos a partir de 2004, o autor passou a se envolver com as questões relacionadas a características de ferramentas e de eficiência de codificação de vídeo. Neste contexto, as principais contribuições do autor, concomitantes às contribuições desta Dissertação, foram participações em debates sobre as características e vantagens da codificação H.264/AVC no momento de escolha do sistema brasileiro, a colaboração técnica e de escrita em partes da norma de codificação de vídeo (N02 Parte 1 ou ABNT NBR 15602-1) do SBTVD junto ao Fórum de TV Digital, e dois artigos referentes à arquitetura do terminal de acesso (receptor) desenvolvido como referência para a TV Digital brasileira (HIRA et al., 2007; CARVALHO et al., 2007).

Neste contexto, em função do recente enfoque de pesquisa do autor em codificação de vídeo, visando possibilitar a evolução da codificação de vídeo do SBTVD, mediante superação das limitações tecnológicas momentâneas identificadas durante a especificação do sistema, surgiu a idéia de tema para esta Dissertação, cujos objetivos propostos foram atingidos com êxito.

6.1 Contribuições

A síntese da principal contribuição do trabalho realizado nesta Dissertação foi a evidência da possibilidade de evolução dos formatos de vídeo no SBTVD através da escalabilidade do H.264/AVC, preservando o legado de equipamentos de recepção já em uso no sistema digital atual.

Além disso, durante o trabalho, o autor motivou-se a investigar de forma mais abrangente o histórico da codificação multimídia (especialmente vídeo), aprofundando-se nos temas multimídia reconfigurável e escalabilidade. O segundo tema resultou no Capítulo 2, que além de ser parte fundamental desta dissertação, constitui uma contribuição como referência para planos futuros de trabalhos relacionados que o LSI pretende desenvolver na área.

No Capítulo 2 foi apresentado o histórico da codificação de vídeo, que incorpora na codificação estado-da-arte do H.264/AVC conceitos que começaram a ser desenvolvidos há mais 60 anos. Foram apresentadas importantes referências que tratam com maior profundidade as ferramentas apresentadas, ou que compilam referências sobre o histórico detalhado da evolução da codificação de vídeo. Além disso, foram apresentados os conceitos de escalabilidade e, principalmente, as características relevantes da codificação H.264/AVC e de sua extensão de escalabilidade (SVC).

As características do SVC mostraram, então, potencial para a evolução dos formatos de vídeo no SBTVD, pela possibilidade de preservação dos receptores fixos legados do sistema, resultantes de investimentos recentes. Para caracterização do receptor legado, foram apresentadas no Capítulo 3 as características relevantes para transmissão de vídeo no SBTVD. Este capítulo registrou também momentos importantes da história da TV Digital no Brasil, sob a perspectiva do autor, que participou diretamente de parte dela.

A partir das características do SVC e das partes relevantes do SBTVD, o autor buscou projetos correlatos para discutir e evidenciar um possível caminho de evolução do SBTVD quanto a formatos de vídeo, apoiado sobre a escalabilidade do SVC. As contribuições sobre o direcionamento considerado encontram-se no Capítulo 4. A viabilidade das considerações apresentadas dependiam não somente de evidências experimentais objetivas, mas também de uma análise subjetiva criteriosa, que não estava no escopo deste trabalho. Desta forma, para que o autor se familiarizasse com equipamentos de transmissão e com a codificação de vídeo SVC, criando referências iniciais para trabalhos futuros, foi especificado e montado um ambiente experimental, também descrito no Capítulo 4.

No ambiente experimental foram encontradas algumas limitações, que demonstram as dificuldades de se trabalhar com equipamentos e ferramentas que implementam funcionalidades no estado-da-arte, pois estes podem estar incompletos, em um estágio preliminar, ou mesmo com falhas de implementação. No entanto, como mostram os resultados do Capítulo 5, foi possível reproduzir alguns testes e avaliações que evidenciaram a validade das premissas utilizadas na codificação escalável. Os resultados evidenciaram também que a escalabilidade representa uma solução vantajosa, frente ao *simulcast*, para a evolução do SBTVD, desde que testes subjetivos, a serem realizados como trabalhos futuros, evidenciem que as taxas praticadas no SBTVD possibilitam qualidade aceitável.

6.2 Trabalhos Futuros

Não foram poucos os trabalhos futuros necessários identificados para viabilizar uma transmissão experimental com um formato superior ao 1080i no SBTVD. Inicialmente, deve ser criado um novo acervo de seqüências de vídeo para testes, composto por uma maior diversidade de seqüências, tanto em termos de características de cenas quanto em taxa de quadros, priorizando as seqüências que tenham sido produzidas nas taxas compatíveis com o SBTVD. São necessárias intervenções no JSVM, software de referência do SVC, tanto para complementar funcionalidades não implementadas, principalmente com relação a controle de taxa e entrelaçamento, quanto para otimizar parâmetros e algoritmos internos, que tornam sua execução muito custosa em processamento, quantidade de memória e tempo, para cada interação de teste.

Identificou-se também a necessidade de trabalho na parte de multiplexação e transmissão para que as ferramentas aceitem o fluxo elementar de vídeo gerado pelo JSVM. Aproveitando este trabalho, podem ser elaboradas extensões que permitam a separação de um fluxo de vídeo escalável entre camadas de transporte distintas, para transmissão por camadas hierárquicas distintas ou por múltiplos canais simultâneos.

A codificação escalável por qualidade não foi considerada necessária nos primeiros cenários de evolução do SBTVD. No entanto, a partir do momento que a taxa do fluxo codificado para um determinado formato exija taxas superiores a de um ou mais canais, este tipo de escalabilidade pode vir a representar uma solução possível para segmentar o fluxo de vídeo em camadas com taxas controladas para flexibilidade de distribuição dos fluxos parciais entre canais (ou até mesmo meios) distintos.

Outro trabalho futuro identificado, que também seria transversal entre as áreas abor-

dadas no Capítulo 3, seria o estudo da relação entre as configurações hierárquicas e de formatação do fluxo de vídeo escalável para otimizar a relação entre potência de transmissão, área de cobertura e diversidade de formatos.

A abordagem dos trabalhos futuros citados pode ser uma grande oportunidade para o Brasil ampliar sua inserção mundial no cenário de desenvolvimento tecnológico, especialmente na área de codificação de vídeo. Uma motivação adicional para isso seria a Copa do Mundo FIFA 2014 de futebol, que será realizada no país. Normalmente, este evento, além de representar uma marca histórica no esporte, também é uma oportunidade de consolidação e demonstração de inovações na produção e distribuição multimídia. No caso, a apresentação de resultados relacionados com a escalabilidade do H.264/AVC no âmbito do SBTVD é uma possibilidade que pode ser considerada.

Referências Bibliográficas

- ABNT. *ABNT NBR 15601: Televisão digital terrestre - Sistema de transmissão*. [S.l.], nov. 2007.
- ABNT. *ABNT NBR 15602-1: Televisão digital terrestre - Codificação de vídeo, áudio e multiplexação. Parte 1: Codificação de vídeo*. [S.l.], nov. 2007.
- ABNT. *ABNT NBR 15602-2: Televisão digital terrestre - Codificação de vídeo, áudio e multiplexação. Parte 2: Codificação de áudio*. [S.l.], nov. 2007.
- ABNT. *ABNT NBR 15602-3: Televisão digital terrestre - Codificação de vídeo, áudio e multiplexação. Parte 3: Sistemas de multiplexação de sinais*. [S.l.], nov. 2007.
- ABNT. *ABNT NBR 15603-1: Televisão digital terrestre - Multiplexação e serviços de informação (SI). Parte 1: Serviços de informação do sistema de radiodifusão*. [S.l.], nov. 2007.
- ABNT. *ABNT NBR 15603-2: Televisão digital terrestre - Multiplexação e serviços de informação (SI). Parte 2: Sintaxes e definições da informação básica de SI*. [S.l.], nov. 2007.
- ABNT. *ABNT NBR 15603-3: Televisão digital terrestre - Multiplexação e serviços de informação (SI). Parte 3: Sintaxe e definições da informação estendida do SI*. [S.l.], nov. 2007.
- BARROS, G. G. *A consistência da interface com o usuário para a TV interativa*. 200 p. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, 158, Trav. 3 - Butantã, São Paulo/SP, Brasil, 2006.
- BRASIL. *Decreto nº 4.901, de 26 de Novembro de 2003*. Brasília, DF, nov. 2003. Institui o projeto do Sistema Brasileiro de TV Digital (SBTVD).
- BRASIL. *Decreto nº 5.820, de 29 de Junho de 2006*. Brasília, DF, jun. 2006. Institui o Sistema Brasileiro de TV Digital Terrestre (SBTVD-T).
- CARVALHO, E. R. et al. The Brazilian digital television system access device architecture. *Journal of Brazilian Computer Society*, v. 13, n. 1, mar. 2007.
- CARVALHO, E. R. de. *Uma plataforma modular para testes com interatividade na TV digital brasileira*. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, 158, Trav. 3 - Butantã, São Paulo/SP, Brasil, 2008.
- CHEN, T. et al. Technologies in mpeg-4 avc frext for improving subjective picture quality in high-definition video compression. In: *International Conference on Consumer Electronics, ICCE, Digest of Technical Papers*. [S.l.: s.n.], 2005. p. 225–226.

- CHOUPANI, R.; WONG, S.; TOLUN, M. *Scalable video coding: a technical report*. [S.l.], set. 2007. 1-33 p.
- CLARKE, R. J. *Digital Compression of Still Images and Video*. Orlando, FL, USA: Academic Press, Inc., 1995. ISBN 012175720X.
- CONSELHO NACIONAL DE ENGENHARIA. *TV Digital: qualidade e interatividade*. Brasília: Instituto Euvaldo Lodi, 2007. 160 p.
- CPQD. *Arquitetura de referência*. Versão AA. PD.30.12.34.0001A/RT-13/AA. Campinas, 2006.
- CPQD. *Especificação técnica de referência*. Versão AA. PD.30.12.34.0001A/RT-14/AA. Campinas, 2006.
- DALPOZ, M. A. S. *Um terminal de acesso digital reconfigurável bidirecional adaptável aos padrões multimídia emergentes*. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, 158, Trav. 3 - Butantã, São Paulo/SP, Brasil, 2005.
- DANYALI, H. *Highly scalable wavelet image and video coding for transmission over heterogeneous networks*. Tese (Doutorado) — School of Electrical, Computer and Telecommunications Engineering at The University of Wollongong, 2004.
- DAUBECHIES, I. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, v. 41, p. 909–996, out. 1988.
- DAVIS, G.; NOSRATINIA, A. Wavelet-based image coding: an overview. *Applied and Computational Control, Signals, and Circuits*, v. 1, n. 1, p. 205–269, 1998.
- EBU. *Current status of high definition television delivery technology*. [S.l.], maio 2008. 26 p.
- FRANÇOIS, E.; VIÉRON, J.; BOTTREAU, V. Interlaced coding in SVC. *IEEE Trans. Circuits Syst. Video Techn.*, v. 17, n. 9, p. 1136–1148, 2007.
- GHANBARI, M. Two-layer coding of video signals for vbr networks. *IEEE Journal on Selected Areas in Communications*, v. 7, p. 771–781, jun. 1989.
- GOODALL, W. Television by pulse code modulation. *Bell System Tech.*, v. 30, p. 33–49, jan. 1951.
- HIRA, C. *ARQUIMEDIA: Uma proposta de arquitetura de software para terminais de acesso à TV digital interativa*. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, 158, Trav. 3 - Butantã, São Paulo/SP, Brasil, 2008.
- HIRA, C. et al. Sistema operacional do terminal de acesso de referência. *Workshop de Sistemas Operacionais*, Rio de Janeiro, RJ, p. 786–795, 2007.
- IBM INSTITUTE FOR BUSINESS VALUE. *The End of television as we know it: a future industry perspective*. New York, USA, 2006. 27 p. Institui o projeto do Sistema Brasileiro de TV Digital (SBTVD).

- IEEE Transactions on circuits and systems for video technology, v. 13, n. 7. 560–762 p.
- ITU-T. *ITU-T H.264: Advanced video coding for generic audiovisual services*. [S.l.], nov. 2007.
- JACK, K. *Video Demystified: A Handbook for the Digital Engineer*. [S.l.]: L L H Technology Publishing, 1995. ISBN 187870723X.
- JUNIOR, G. B. *Sintonizador-demodulador para o Sistema Brasileiro de TV Digital*. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, 158, Trav. 3 - Butantã, São Paulo/SP, Brasil, 2008.
- JVT. *SVC Verification Test Report - N95777*. [S.l.], jan. 2007. 19 p. Disponível em: http://www.chiariglione.org/mpeg/working_documents/mpeg-04/svc/svc-vt.zip. Acesso em: 10 jan. 2009.
- KOUADIO, A. et al. *SVC - a highly-scalable version of the H.264/VC*. [S.l.], maio 2008. 22 p.
- LI, W. Scalable video coding with fine granularity scalability. *Consumer Electronics, 1999. ICCE. International Conference on*, Los Angeles, CA, USA, p. 306–307, 1999.
- LUTHRA, A.; SULLIVAN, G.; WIEGAND, T. Introduction to the special issue on the H.264/AVC video coding standard. In: *IEEE Transactions on Circuits and Systems for Video Technology*. [S.l.: s.n.], 2003. v. 13, n. 7, p. 557–559.
- OHM, J.-R. Advances in scalable video coding. In: *Proceedings of the IEEE*. [S.l.: s.n.], 2005. v. 93, n. 1, p. 42–56.
- PEREIRA, F. A revolução multimídia: o papel da normalização. jun. 1997.
- POYNTON, C. *Digital Video and HDTV Algorithms and Interfaces*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003. ISBN 1558607927.
- PROCEEDINGS of the IEEE, v. 93, n. 1. 1–208 p.
- READER, C. History of MPEG video compression. Ver. 4.0. Joint Video Team. out. 2002.
- SAKAIDA, S. et al. The super hi-vision codec. *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, set. 2007.
- SCHWARZ, H. et al. Technical description of the HHI proposal for SVC CE1. Doc. M11244. out. 2004.
- SCHWARZ, H.; MARPE, D.; WIEGAND, T. Hierarchical B pictures. Doc. JVT-P014. jul. 2005.
- SCHWARZ, H.; MARPE, D.; WIEGAND, T. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Trans. Circuits Syst. Video Techn.*, v. 17, n. 9, p. 1103–1120, 2007.
- SHISHIKUI, Y.; FUJITA, Y.; KUBOTA, K. *Super hi-vision - the star of the show*. [S.l.], jan. 2009. 13 p.

- SUGAWARA, M. *Super hi-vision - research on a future ultra-HDTV system*. [S.l.], jun. 2008. 7 p.
- SUGAWARA, M. et al. Research on human factors in ultra-high-definition television to determine its specification. *SMPTE Motion Imaging Journal*, abr. 2008.
- SULLIVAN, G. J.; WIEGAND, T. Video compression - from concepts to the H.264/AVC standard. In: *Proceedings of the IEEE*. [S.l.: s.n.], 2005. v. 93, n. 1, p. 18–31.
- UEHARA, M. Application of mpeg-2 systems to terrestrial isdb (isdb-t). *Proceedings of the IEEE*, v. 94, n. 1, p. 261–268, jan. 2006. ISSN 0018-9219.
- VAZ, R. A. *Uma interface de comunicação sem fio em TV digital baseada em rádio definido por programa de computador*. 200 p. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, Av. Prof. Luciano Gualberto, 158, Trav. 3 - Butantã, São Paulo/SP, Brasil, 2007.
- VQEG HDTV GROUP. *Test plan for evaluation of video quality models for use with High Definition TV content*. [S.l.], ago. 2008. 40 p. Disponível em: ftp://vqeg.its.bldrdoc.gov/Documents/Projects/hdtv/other/VQEG_HDTV_testplan_v2_2_proposals.doc. Acesso em: 02 dez. 2008.
- WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, v. 13, p. 600–612, 2004.
- WIEGAND, T. et al. Overview of the H.264/AVC video coding standard. In: *IEEE Transactions on Circuits and Systems for Video Technology*. [S.l.: s.n.], 2003. v. 13, n. 7, p. 560–576.
- WIEGAND, T. et al. Introduction to the special issue on scalable video coding-standardization and beyond. *IEEE Trans. Circuits Syst. Video Techn.*, v. 17, n. 9, p. 1099–1102, 2007.
- WU, F. et al. Efficient and universal scalable video coding. In: *Proceedings of the IEEE*. [S.l.: s.n.], 2002. v. 2, p. II–37–II–40.
- WU, H. R.; RAO, K. R. *Digital video image quality and perceptual coding (Signal processing and communications)*. Boca Raton, FL, USA: CRC Press, Inc., 2005. ISBN 0824727770.

Apêndice A – Representação de Vídeo

Este Apêndice constitui uma introdução básica sobre a representação de vídeo amplamente utilizada para conteúdo em duas dimensões. Os conceitos apresentados podem ser aprofundados em Poynton (2003) e Jack (1995).

Desde os estudos e experimentos anteriores à demonstração de Baird descrita no Capítulo 1, a captura de informação visual para transmissão ou armazenamento analógico está baseada na sistemática apresentada na Figura A.1. Por esta sistemática, um foto sensor varre uma área retangular capturada por uma lente da esquerda para a direita e de cima para baixo, transformando então a informação de uma imagem estática de duas dimensões em um sinal elétrico unidimensional, variando no tempo. Repetindo-se o processo de captura de imagem em intervalos de tempo constantes, tem-se a informação de movimento. O sinal elétrico completo resultante carrega, dessa forma, as informações de varreduras consecutivas de um retângulo de cena.

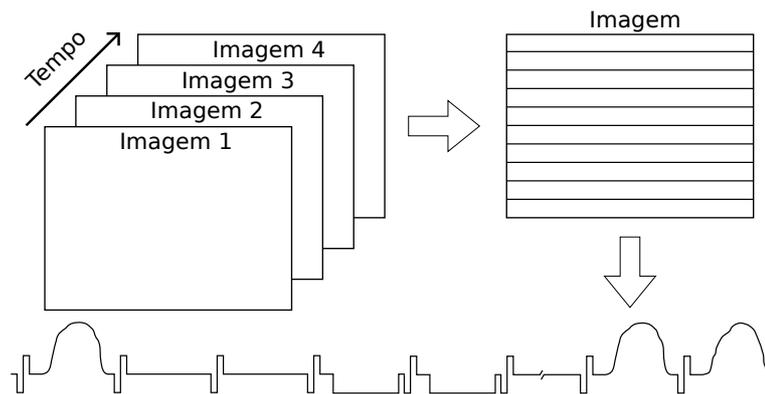


Figura A.1: Sistemática de captura de vídeo analógico

A qualidade do vídeo é determinada tanto pela qualidade de cada imagem independente quanto pela taxa de troca de imagens.

A qualidade da imagem estática por sua vez é determinada tanto pelo número de linhas de varredura quanto pela banda passante do sinal gerado, que determina a capacidade

de representar variações de informação em uma mesma linha. Em outras palavras, de acordo com a banda e o número de linhas, maior ou menor é a riqueza de detalhes de cada imagem.

Além da informação visual propriamente dita, um sinal de vídeo analógico carrega ainda pulsos de sinalização de sincronismo, que indicam tanto os pontos de separação entre imagens quanto os pontos de quebra de linha de cada imagem. Esses pulsos são conhecidos como pulsos de sincronismo ou retraço horizontal, quando sinalizam quebra de linha, e pulsos de sincronismo ou retraço vertical, quando sinalizam separação entre imagens.

Quando a taxa de troca de imagens em um vídeo (ou simplesmente taxa do vídeo) não é suficientemente elevada, pode-se perceber cintilação das imagens ou descontinuidade em regiões de cenas onde há movimento. A taxa mínima necessária para evitar que essas características indesejáveis tornem-se perceptíveis depende principalmente do espectador e da intensidade de iluminação e do tamanho da tela onde o vídeo está sendo exibido. Como regra geral, quanto mais luminosa a fonte de vídeo e quanto maior o tamanho da tela, maior deve ser a taxa do vídeo. Há um limite prático, porém, em função das características da visão humana, de forma que a taxa não cresce indiscriminadamente com esses fatores. Verifica-se que sistemas com até 80 Hz de taxa de vídeo são suficientes para telas grandes, bem iluminadas e que cubram a visão periférica da grande maioria dos espectadores, onde a sensibilidade à cintilação e a movimentos é maior.

A.1 Entrelaçamento

Nos sistemas televisivos que surgiram na década de 30, tanto pelas características da visão humana de resposta à frequência de vídeo quanto por limitações dos tubos de raios catódicos, fixou-se a taxa de vídeo de acordo com a corrente alternada das redes elétricas em 50 ou 60 Hz, dependendo do país. Com esses valores, evitavam-se interferências visuais do circuito de alimentação dos aparelhos televisores, mantendo-se a taxa de vídeo ainda na faixa adequada para a visão em ambientes com boa iluminação. Para as frequências de 50 ou 60 Hz, no entanto, por limitações tecnológicas dos tubos de raios, a resolução máxima dos primeiros aparelhos era de pouco mais de 200 linhas horizontais.

Essa limitação era indesejável, e logo a necessidade de melhoria de qualidade de vídeo exigiria aumento de resolução nas transmissões. Mas como os parâmetros de frequência e resolução de vídeo são os principais na determinação da banda necessária para transmissão

de sinal analógico, e como o aumento de banda poderia inviabilizar o sistema, não se desejava aumentar a banda de transmissão e tão pouco diminuir a frequência do vídeo. A técnica de entrelaçamento, patenteada na década de 30 por Randall C. Ballard, resolveu por muitos anos essa questão aparentemente contraditória.

A técnica, ainda amplamente utilizada, consiste em varrer, conforme mostra a Figura A.2, alternadamente somente as linhas superiores ou inferiores entre imagens consecutivas, considerando-se as linhas duas a duas. Uma imagem completa normalmente é referenciada na literatura como quadro de vídeo, enquanto que as imagens compostas somente pelas linhas superiores ou inferiores são referenciadas respectivamente por campo superior ou campo inferior de vídeo.

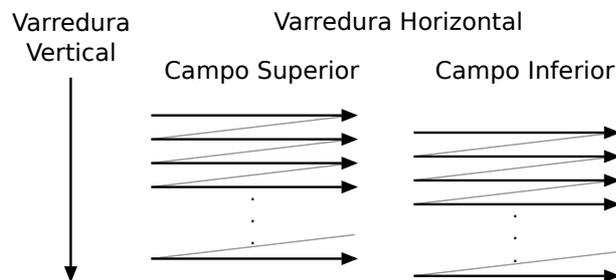


Figura A.2: Entrelaçamento do sinal de vídeo

Com o entrelaçamento, é possível realizar a captura de duas vezes mais campos do que quadros num determinado intervalo de tempo, obtendo-se a mesma quantidade de informação, pois cada campo apresenta a metade da quantidade de linhas de um quadro completo. Analogamente, mantendo-se a taxa do vídeo e a quantidade de informação tratável pelo sistema, outra opção possibilitada pelo entrelaçamento é a captura de campos com a mesma quantidade de linhas dos quadros originais, de forma que dois campos consecutivos somados resultem num quadro com o dobro de linhas do quadro original, aumentando a resolução final do vídeo.

Para a transmissão televisiva, como a frequência prevista nos sistemas já era suficiente para a visão humana, o entrelaçamento foi utilizado para duplicar a quantidade de linhas transmitidas por quadro de vídeo, separando cada quadro em dois campos consecutivos. Ou seja, o vídeo para televisão passou a ser transmitido como uma seqüência de campos a 50 ou 60 Hz com, respectivamente, 626 ou 525 linhas, aumentando a resolução do vídeo.

A partir da década de 70, os computadores voltaram a enfatizar a necessidade de monitores progressivos (sem entrelaçamento), pois trabalhavam fundamentalmente com imagens artificiais estáticas, ou seja, sem movimento. Ainda assim, para a televisão, vídeo entrelaçado continuou sendo a melhor alternativa, pois possibilitava sempre maior

resolução espacial para o mesmo nível de complexidade sistêmica.

Atualmente, a resolução dos sistemas de televisão é tipicamente de 1080 linhas e, embora a transmissão seja digital, existem interfaces analógicas para essa resolução.

A.2 Composição do Sinal de Vídeo

Os primeiros sistemas de captura de vídeo geravam um único sinal que representava a quantidade de luminosidade (sinal de luma) das imagens ao longo da varredura. Por isso eram sistemas conhecidos como preto-e-branco¹, pois não possuíam informações sobre as cores das imagens.

Sistemas posteriores passaram a gerar, durante a varredura, os três sinais das cores primárias vermelho, verde e azul, conhecidos como $R'G'B'$ (do inglês *Red, Green, Blue*). No entanto, para compatibilidade com os sistemas preto-e-branco e pela alta correlação entre os três sinais das cores primárias, optou-se por buscar uma transformação capaz de converter os sinais $R'G'B'$ em outros três sinais que ficaram conhecidos como $Y'P_BP_R$ ² pois um deles era equivalente ao sinal de luma (Y') dos sistemas preto-e-branco e os outros dois sinais continham informações complementares sobre as componentes de cores (conhecidos como sinais de croma, P_B representa o complemento de azul e P_R o complemento de vermelho).

A transformação de $R'G'B'$ para $Y'P_BP_R$ tornou-se ainda mais conveniente, pois a visão do ser humano é menos sensível às informações dos sinais P_B e P_R do que do sinal Y' . Essa característica pôde também ser explorada para diminuir a quantidade de informação de um vídeo, pois para uma determinada quantidade de informação de luma, é necessário menos informação para os sinais de croma correspondentes.

Este processo de redução da quantidade de informação de croma, conhecido como subamostragem de cor, continua sendo explorado nos sistemas digitais modernos.

A representação analógica de sinais de vídeo, no entanto, não se limitou a representar as informações em sinais de banda base ou apenas em sinais separados para luma e croma.

¹Um nome mais adequado para os sistemas preto-e-branco teria sido sistemas de tons de cinza, pois representam também outros tons entre estes dois extremos.

²Assume-se neste texto que os sinais de vídeo usados para transmissão e armazenamento, sejam as componentes primárias ou os sinais derivados, estão já corrigidos pelo fator gama, e são valores que quantificam intensidades segundo definições de alguma recomendação internacional, como ITU-R Rec. BT.601 ou ITU-R Rec. BT.709, que determinam a relação das intensidades dos sinais e as cores representadas com base em diagrama de cromaticidade. Informação detalhada a respeito do tema pode ser obtida em Poynton (2003).

Mesmo os sinais em preto-e-branco já eram modulados e as técnicas de modulação foram então aprimoradas para acomodar os sinais de cores quando estes surgiram.

Dessa forma, as principais representações criadas para sinais analógicos de vídeo foram as componentes analógicas $Y'P_BP_R$, $R'G'B'$, $Y'UV$ (representadas por sinais independentes) e os sinais compostos de vídeo NTSC, PAL e SECAM (representados por sinais modulados).

A.3 Representações Digitais

Na busca por representações mais eficientes para um sinal de vídeo, uma primeira abordagem já nos anos 50, que inclusive marca o início das pesquisas em codificação de vídeo digital com o trabalho de Goodall (1951), conforme reconhecimento de Clarke (1995), consistia na digitalização simples por PCM do sinal analógico de vídeo. Esta abordagem resultou na técnica de amostragem temporal e quantização dos sinais de luma e croma, representados no contexto digital por $Y'C_BC_R$ ³. A amostragem PCM forma uma seqüência de matrizes de pontos, conhecidos como *pels* ou *pixels*⁴ que representam as componentes de cada imagem de um vídeo progressivo ou entrelaçado. Dessa forma, cada imagem é tipicamente composta pela sobreposição de três matrizes bi-dimensionais, formadas por amostras com valores inteiros.

O resultado da amostragem simples, no entanto, era muito pouco eficiente, ainda que como nas representações analógicas, fosse possível subamostrar os sinais de croma. Isso porque a informação visual geralmente é altamente redundante e parte dela irrelevante, ou seja, além da informação redundante que se pode omitir, mas que se pode recuperar na íntegra por processamento de algoritmos matemáticos, ainda há informação que o cérebro humano não é capaz de assimilar. Essa informação pode ser descartada sem comprometer a interpretação subjetiva da imagem. O resultado da amostragem simples, no entanto, era muito pouco eficiente, ainda que como nas representações analógicas, fosse possível subamostrar os sinais de croma. Isso porque a informação visual geralmente é altamente redundante e parte dela irrelevante, ou seja, além da informação redundante que se pode omitir, mas que se pode recuperar na íntegra por processamento de algoritmos matemáticos, ainda há informação que o cérebro humano não é capaz de assimilar. Essa informação pode ser descartada sem comprometer a interpretação subjetiva da imagem.

Estas características já haviam motivado a comunidade científica a investigar técnicas

³ C_B : equivalente ao P_B e C_R : equivalente ao P_R

⁴Os *pels* ou *pixels* representam elementos de figura, que em inglês são os *Picture Elements*.

para eliminação das informações redundantes ou irrelevantes dos sinais de vídeo, o que deu origem ao campo de investigação que ficou largamente conhecido por compressão e codificação de imagem e vídeo.

Apêndice B – Códigos-Fontes

Neste apêndice são apresentados os códigos-fontes dos roteiros (*scripts*) utilizados no ambiente de investigação experimental para preparação de conteúdo, realização de testes ou consolidação de resultados. São apresentados também arquivos de configuração quando necessários em conjunto com os roteiros.

Definido um diretório raiz para o ambiente experimental (identificado por `<env_root>`), a estrutura interna deste diretório deve respeitar uma hierarquia. Antes de cada arquivo, é indicado em qual sistema operacional o mesmo deve/pode ser executado e em que lugar da hierarquia ele deve ser armazenado. O texto `<env_root>` deve ser substituído pelo caminho completo do diretório raiz quando aparecer em um roteiro.

Num primeiro momento, a forma mais simples de reproduzir todos os experimentos é manter uma réplica completa da estrutura do ambiente tanto em uma máquina Linux quanto em uma máquina Microsoft WindowsTM com Cygwin instalado. À medida que os experimentos são realizados em cada ambiente, os resultados devem ser propagados para o outro. É possível também compartilhar o diretório `<env_root>` entre os ambientes por rede, mas os testes tornam-se ainda mais demorados.

Todos os roteiros devem ser executados a partir de `<env_root>/scripts/`. Espera-se que os aplicativos MPlayer/MEncoder estejam instalados e possam ser executados de qualquer diretório do sistema em uso.

B.1 Roteiro de Configuração de Variáveis

Este roteiro é, na verdade, um arquivo que define valores de variáveis utilizadas pelos outros roteiros, a fim de facilitar a adaptação entre ambientes.

(Linux/Windows): `<env_root>/scripts/config.sh`

```

1 #Experimental Environment Root Directory
2 ENV_ROOT="<env_root>"

```

```

3
4 #Path to the JSVM binaries
5 JSVM_PATH="$ENV_ROOT/Software/jsvm/bin"
6
7 #Path to Windows Media Player 6.x binary
8 MPLAYER6='cygpath -u \'$PROGRAMFILES/Windows Media Player/mplayer2.exe\'"'
9
10 #For the original content
11 VQEG_SGL_ROOT="$ENV_ROOT"
12 VQEG_SGL_SEQS="$VQEG_SGL_ROOT/vqeg.its.blrdoc.gov/HDTV/SVT_MultiFormat"
13 VQEG_YUV_SEQS="$ENV_ROOT/Videos/yuv/svt"
14
15 #For the clipped seqs
16 VQEG_YUV_CLIPPED_SEQS="$ENV_ROOT/Videos/yuv/clipped/sinc"
17 DOWNCONV_YUV_CLIPPED_SEQS="$ENV_ROOT/Videos/yuv/clipped/down"
18 SKIP_FRAMES=50
19 FRAMES=100
20
21 #For the avi seqs
22 AVLSTREAMS_SRC="$VQEG_YUV_CLIPPED_SEQS"
23 AVLSTREAMS="$ENV_ROOT/Videos/avi"
24
25 #For the tests
26 TESTS="$ENV_ROOT/Videos/tests"
27 TMP="/tmp"

```

B.2 Roteiros de Preparação das Seqüências de Vídeo

A preparação das seqüências envolve copiá-las do servidor, gerar seqüências I420 a partir dos arquivos sgi16 e truncar as seqüências para diminuir o tempo dos testes e o tamanho dos arquivos.

B.2.1 Cópia dos Arquivos Originais do Servidor

O roteiro a seguir baixa (realiza o *download*) as seqüências CrowdRun, ParkJoy, DucksTakeOff, InToTree e OldTownCroos do servidor da SVT, armazenando-as no local indicado pela variável VQEG_SGL_ROOT (ver Sessão B.1), que deve estar previamente criado e vazio. O processo de *download* é executado em segundo plano, pois geralmente leva dias para terminar.

(Linux): <env_root>/scripts/downloadSgiSeqs.sh

```

1 #!/bin/bash
2 source config.sh
3
4 pushd $VQEG_SGL_ROOT
5 wget -t0 -b -q -c -x -r ftp://vqeg.its.blrdoc.gov/HDTV/SVT_MultiFormat/
6 popd

```

B.2.2 Conversão de Formato Sgi16 em I420

Este roteiro realiza, com a ferramenta ImageMagick a conversão das seqüências em sgi16 baixadas no diretório apontado por VQEG_SGI_ROOT para seqüências I420, gravadas no diretório apontado por VQEG_YUV_SEQS.

(Linux): <env_root>/scripts/genYUVSeqs.sh

```

1 #!/bin/bash
2 source config.sh
3
4 echo Generating complete SVT Sinc based YUV Streams
5 for i in `find $VQEG_SGI_SEQS/* -mindepth 1 -maxdepth 1 -type d`; do
6   DEST_DIR="$VQEG_YUV_SEQS/`basename \"$i\"`"
7   DEST_FILE="$DEST_DIR/`basename \"$i\"`.yuv"
8   if [ ! -e "$DEST_FILE" ]; then
9     echo Generating $DEST_FILE
10    rm -rf /tmp/yuv_tmp
11    mkdir /tmp/yuv_tmp
12    echo "Reading directory: $i"
13    for j in `ls $i`; do
14      echo "  Processing: $j"
15      convert $i/$j /tmp/yuv_tmp/`echo $j | sed 's/ /%/'`.sgi`.yuv
16    done
17    mkdir -p "$DEST_DIR"
18    cat /tmp/yuv_tmp/* > $DEST_FILE
19  else
20    echo "File $DEST_FILE already generated"
21  fi
22 done
23 rm -rf /tmp/yuv_tmp

```

B.2.3 Truncamento das Seqüências

Este roteiro trunca as seqüência geradas pelo roteiro da Sessão B.2.2 em quantidade de quadros dada por FRAMES, pulando os primeiros quadros determinados por SKIP_FRAMES. As seqüências truncadas são geradas em VQEG_YUV_CLIPPED_SEQS, e uma segunda versão das seqüências truncadas com a ferramenta de subamostragem do JSVM é gerada em DOWNCONV_YUV_CLIPPED_SEQS. Para economizar espaço em disco, as seqüências das sessões anteriores podem ser descartadas.

(Linux): <env_root>/scripts/genClippedYUVSeqs.sh

```

1 #!/bin/bash
2 source config.sh
3
4 echo "Generating SVT Sinc Based Clipped YUV Seqs"
5 for i in `find $VQEG_YUV_SEQS/* -type d`; do
6   DEST_DIR="$VQEG_YUV_CLIPPED_SEQS/${i#$VQEG_YUV_SEQS}clipped_"
7   for j in `ls $i`; do

```

```

8     SRC_FILE=${i}/${j}
9     if echo "$SRC_FILE" | grep -q "TestPatterns" ; then
10        echo "Ignoring $SRC_FILE"
11        continue
12    fi
13    DEST_FILE=$DEST_DIR/${j%.yuv}clipped.yuv
14    if [ ! -e $DEST_FILE ] ; then
15        mkdir -p $DEST_DIR
16        echo "Generating $FRAMES frames; skiping $SKIP_FRAMES."
17        echo "From: $SRC_FILE"
18        echo "To: $DEST_FILE"
19        if echo "$j" | grep -q "576i" ; then
20            FRAME_SIZE="720*576*1.5/512/2"
21            dd bs=512 skip='echo $SKIP_FRAMES*$FRAME_SIZE | bc' count='echo
22                $FRAMES*$FRAME_SIZE | bc' if=$SRC_FILE of=$DEST_FILE
23        elif echo "$j" | grep -q "720p" ; then
24            FRAME_SIZE="1280*720*1.5/2048"
25            dd bs=2048 skip='echo $SKIP_FRAMES*$FRAME_SIZE | bc' count='echo
26                $FRAMES*$FRAME_SIZE | bc' if=$SRC_FILE of=$DEST_FILE
27        elif echo "$j" | grep -q "1080i" ; then
28            FRAME_SIZE="1920*1080*1.5/512/2"
29            dd bs=512 skip='echo $SKIP_FRAMES*$FRAME_SIZE | bc' count='echo
30                $FRAMES*$FRAME_SIZE | bc' if=$SRC_FILE of=$DEST_FILE
31        elif echo "$j" | grep -q "1080p" ; then
32            FRAME_SIZE="1920*1080*1.5/512"
33            dd bs=512 skip='echo $SKIP_FRAMES*$FRAME_SIZE | bc' count='echo
34                $FRAMES*$FRAME_SIZE | bc' if=$SRC_FILE of=$DEST_FILE
35        else
36            echo "FORMAT NOT SUPPORTED!"
37        fi
38        echo ""
39    else
40        echo "File $DEST_FILE already generated"
41    fi
42 done
43
44 echo "Generating JSVM DownConv Based Clipped YUV Seqs"
45 SRC_DIR="find $VQEG_YUV_CLIPPED_SEQS -type d -regex '.*Master.*'"
46 MASTER_RES="echo $SRC_DIR | sed 's:.*^\([0-9]*[ip][0-9]*\)_.*:1:'"
47 for i in `find $VQEG_YUV_CLIPPED_SEQS/* -type d -regex '.*SINC.*'` ; do
48     RES="echo $i | sed 's:.*^\([0-9]*[ip][0-9]*\)_\^[^_]*_.*:1:'"
49     DEST_DIR="echo $i | sed s:$VQEG_YUV_CLIPPED_SEQS:
50         $DOWNCONV_YUV_CLIPPED_SEQS: | sed s:SINC_FILTER:JSVM_DOWNCONV:"
51     for j in `find $SRC_DIR -type f` ; do
52         DEST_FILE="echo $j | sed s:$SRC_DIR:$DEST_DIR: | sed s/$MASTER_RES/$RES/ | sed s
53             :MASTER:JSVM_DOWNCONV:"
54         if [ ! -e $DEST_FILE ] ; then
55             mkdir -p $DEST_DIR
56             echo "Generating $FRAMES frames"
57             echo "From: $j"
58             echo "To: $DEST_FILE"
59             if echo "$i" | grep -q "576i" ; then
60                 "$JSVM_PATH/DownConvertStatic" 3840 2160 "$j" 720 576 "$DEST_FILE" -

```

```

        resample_mode 4
59     elif echo "$i" | grep -q "720p" ; then
60         "$JSVM_PATH/DownConvertStatic" 3840 2160 "$j" 1280 720 "$DEST_FILE"
61     elif echo "$i" | grep -q "1080i" ; then
62         "$JSVM_PATH/DownConvertStatic" 3840 2160 "$j" 1920 1080 "$DEST_FILE" -
        resample_mode 4
63     elif echo "$i" | grep -q "1080p" ; then
64         "$JSVM_PATH/DownConvertStatic" 3840 2160 "$j" 1920 1080 "$DEST_FILE"
65     fi
66     echo ""
67 fi
68 done
69 done

```

B.2.4 Encapsulamento de Seqüências Truncadas em Camada de Transporte AVI

Este roteiro encapsula as seqüências truncadas apontadas por AVI_STREAMS_SRC (somente as seqüências baseadas no ImageMagic nesta versão) em um fluxo AVI com codificação H.264/AVC sem perdas, para diminuir o tamanho dos arquivos finais sem comprometer a qualidade. As seqüências AVI são armazenadas em AVI_STREAMS (somente progressivas nesta versão).

(Linux): <env_root>/scripts/genClippedAVISeqs.sh

```

1  #!/bin/bash
2  source config.sh
3
4  echo "Generating MPlayer Based AVI Streams"
5  for i in `find $AVI_STREAMS_SRC/* -type d -regex '.*p50.*'` ; do
6      DEST_DIR=`echo $i | sed s:$AVI_STREAMS_SRC:$AVI_STREAMS: | sed s/p50/p5994/`
7      for j in `ls $i` ; do
8          SRC_FILE="$i/$j"
9          DEST_FILE="$DEST_DIR/echo $j | sed s/\.yuv/.avi/ | sed s/p50/p5994/"
10         if [ ! -e $DEST_FILE ] ; then
11             mkdir -p $DEST_DIR
12             echo "Generating $FRAMES frames"
13             echo "From: $SRC_FILE"
14             echo "To: $DEST_FILE"
15             if echo "$j" | grep -q "720p" ; then
16                 mencoder $SRC_FILE -demuxer rawvideo -rawvideo w=1280:h=720:fps
                     =60000/1001:format=i420 -ovc x264 -x264encopts qp=0 -ofps 60000/1001 -of
                     avi -o $DEST_FILE
17             elif echo "$j" | grep -q "1080p" ; then
18                 mencoder $SRC_FILE -demuxer rawvideo -rawvideo w=1920:h=1080:fps
                     =60000/1001:format=i420 -ovc x264 -x264encopts qp=0 -ofps 60000/1001 -of
                     avi -o $DEST_FILE
19             elif echo "$j" | grep -q "2160p" ; then
20                 #mencoder $SRC_FILE -demuxer rawvideo -rawvideo w=3840:h=2160:fps
                     =60000/1001:format=i420 -ovc x264 -x264encopts qp=0 -ofps 60000/1001 -of
                     avi -o $DEST_FILE
21             echo "Not supported by mencoder..."

```

```

22         else
23             echo "FORMAT NOT SUPPORTED!"
24         fi
25         echo ""
26     else
27         echo "File $DEST_FILE already generated"
28     fi
29 done
30 find "$AVI_STREAMS" -empty | xargs rm -rf
31 done

```

B.3 Visualização das Seqüências I420

Este roteiro deve ser chamado com os seguintes parâmetros de entrada, na seqüência apresentada: `arquivo_yuv`, `largura_do_vídeo`, `altura_do_vídeo`, `taxa_de_quadros`. Serve para visualizar qualquer arquivo com terminação `.yuv` do ambiente experimental.

(Linux): `<env_root>/scripts/playYUVSeq.sh`

```

1 #!/bin/bash
2 if [ $1 ]
3 then
4     mplayer $1 -demuxer rawvideo -rawvideo w=$2:h=$3:fps=$4:format=i420 -vo sdl
5 else
6     echo "$0 <YUV_FILE> <WIDTH> <HEIGHT> <FRAME_RATE>"
7 fi

```

B.4 Comparação entre Estruturas Tradicionais e Estrutura HBF

B.4.1 Configuração do Codificador AteMe

Este arquivo serve de base para a geração de todos os arquivos de configuração (120 combinações) com variações de parâmetros para a codificação das seqüências AVI no software da AteMe.

(Windows): `<env_root>/scripts/tests/StructuresComparison/template.xjb`

```

1 <?xml version='1.0' encoding='UTF-8' ?>
2 <job version='1.0.1'>
3     <time>2009-01-19 22:34:46</time>
4     <designer>ATEME KFE Designer version 2.0.1.1</designer>
5     <name><JOB_NAME></name>
6     <state>waiting</state>
7     <priority>normal</priority>
8     <input index='1'>

```

```

 9      <name></name>
10    </input>
11    < ifile index='1' input_index='1'>
12      < file><AVI_FILE></file>
13    </ ifile >
14    <iservice index='10001' input_index='1'>
15      <id>0</id>
16      <name></name>
17    </iservice>
18    <ivstream index='65536' service_index='10001'>
19      <pid>0</pid>
20      <codec>avc</codec>
21      <color_space>invalid</color_space>
22      <width><WIDTH></width>
23      <height><HEIGHT></height>
24      <sample_rate_num>60000</sample_rate_num>
25      <sample_rate_den>1001</sample_rate_den>
26      <length>0</length>
27      <field_order>topFieldFirst</field_order>
28      <dar_num>16</dar_num>
29      <dar_den>9</dar_den>
30      <bit_rate>0</bit_rate>
31    </ivstream>
32    <vtrack index='10001' stream_index='65536' service_index='10001'>
33      <field_order>default</field_order>
34      <dar_num>0</dar_num>
35      <dar_den>0</dar_den>
36    </vtrack>
37    <vtrack_avc index='10001'>
38      <profile>high</profile>
39      <level>l40</level>
40      <rate_control>cbr</rate_control>
41      <bit_rate><BITRATE></bit_rate>
42      <bit_rate_min>1000000</bit_rate_min>
43      <bit_rate_max>20000000</bit_rate_max>
44      <quality>extra</quality>
45      <entropy>cabac</entropy>
46      <interlacing>progressive</interlacing>
47      <cpb_size>1000</cpb_size>
48      <cpb_unit>ms</cpb_unit>
49      <gop_size_min>25</gop_size_min>
50      <gop_size_max>25</gop_size_max>
51      <close_gops />
52      <idr_strategy>always</idr_strategy>
53      <reference_count>2</reference_count>
54      <deblocking_strength>0</deblocking_strength>
55      <bframe_min><BFRAMES></bframe_min>
56      <bframe_max><BFRAMES></bframe_max>
57      <use_hierarchical_bframes />
58      <use_psycho_visual />
59      <slice_count>0</slice_count>
60      <use_hrd />
61      <colorimetry>bt709</colorimetry>
62      <format>unspecified</format>
63      <use_weighted_prediction />
64    </vtrack_avc>
65    <mts index='1'>
66      <id>0</id>

```

```

67     <bit_rate_min>0</bit_rate_min>
68     <pat_period>100</pat_period>
69     <program_number>1</program_number>
70     <pmt_pid>50</pmt_pid>
71     <pmt_period>100</pmt_period>
72     <vbi_pid>0</vbi_pid>
73 </mts>
74 <mts_vtrack index='10001'>
75     <pid>100</pid>
76 </mts_vtrack>
77 <ofile index='1'>
78     <file><TS_FILE></file>
79     <write_strategy>overwrite</write_strategy>
80 </ofile>
81 </job>

```

B.4.2 Geração dos Arquivos de Configuração para o Codificador AteMe

Este roteiro gera, a partir do arquivo da Sessão B.4.1, os arquivos de configuração no diretório `<env_root>/tests/StructuresComparison/jobs`, que devem ser carregados manualmente no software da AteMe para realizar as codificações. Deve-se ter cuidado com as permissões dos diretórios criados automaticamente no Cygwin, pois o software da AteMe pode não conseguir acessar os arquivos de configuração ou criar os fluxos codificados, que são armazenados em `<env_root>/tests/StructuresComparison/ts`.

(Windows): `<env_root>/scripts/tests/StructuresComparison/genCodingJobs.sh`

```

1  #!/bin/bash
2  #windows
3  source config.sh
4  TEST="$TESTS/StructuresComparison"
5  TEST_SCRIPTS="$PWD/$(dirname $0)"
6
7  CODING_JOBS="$TEST/jobs"
8  TS_STREAMS="$TEST/ts"
9
10 structures=(HBF IBBP IBBCP)
11 bitrates=(4000000 8000000 12000000 16000000)
12
13 echo "Generating Structures Comparison Coding Jobs for ATEME KFE"
14 for BITRATE in ${bitrates[@]} ; do
15     for i in `find $AVL_STREAMS/* -type d` ; do
16         DEST_DIR="$CODING_JOBS/((BITRATE/1000000))Mbps/${i#$AVL_STREAMS}"
17         TS_DIR="$TS_STREAMS/((BITRATE/1000000))Mbps/${i#$AVL_STREAMS}"
18         mkdir -p $DEST_DIR
19         mkdir -p $TS_DIR
20         for j in `ls $i` ; do
21             SRC_FILE="$i/$j"
22             for k in ${structures[@]} ; do
23                 JOB_NAME="$j%.avi${(BITRATE/1000000)}Mbps-${k}-"

```

```

24 TS_FILE="$TS_DIR/${JOB_NAME}.ts"
25 TEMPLATE="$TEST_SCRIPTS}/template.xjb"
26 DEST_FILE="$DEST_DIR/${JOB_NAME}.xjb"
27 if [ ! -e $DEST_FILE ] ; then
28     echo "Generating coding job:"
29     echo "$DEST_FILE"
30     if echo "$j" | grep -q "720p" ; then
31         WIDTH="1280"; HEIGHT="720"
32     elif echo "$j" | grep -q "1080p" ; then
33         WIDTH="1920"; HEIGHT="1080"
34     else
35         continue
36     fi
37 BFRAMES=3; unset USEHBF
38 if [ "$k" == "IBBP" ] ; then
39     BFRAMES=2
40 elif [ "$k" == "HBF" ] ; then
41     USEHBF=1
42     fi
43 SRC_FILE="`cygpath -m $SRC_FILE`"
44 TS_FILE="`cygpath -m $TS_FILE`"
45 cat $TEMPLATE | sed -e s,'<JOB_NAME>','$JOB_NAME', -e s,'<
    AVI_FILE>','$SRC_FILE', -e s,'<TS_FILE>','$TS_FILE', -e s,'<
    WIDTH>','$WIDTH', -e s,'<HEIGHT>','$HEIGHT', -e s,'<BFRAMES
    >','$BFRAMES', -e s,'<BITRATE>','$BITRATE', > ${DEST_FILE}.
    tmp
46 if [ ! $USEHBF ] ; then
47     cat ${DEST_FILE}.tmp | sed '/use.hierarchical.bframes/d' > $DEST_FILE
48     ; rm ${DEST_FILE}.tmp
49 else
50     mv ${DEST_FILE}.tmp $DEST_FILE
51     fi
52     echo ""
53     fi
54     done
55     done
56     done

```

B.4.3 Geração das Tabelas de PSNR

Para gerar as tabelas dos valores de qualidade PSNR das seqüências codificadas com o software da Ate me é necessário primeiro decodificá-las para poder compará-las quadro a quadro com as seqüências originais. Isso é feito em um diretório temporário pelo roteiro a seguir, que armazena as tabelas com os resultados finais em `<env_root>/tests/Structures-Comparison/psnr`.

(Windows): `<env_root>/scripts/tests/StructuresComparison/genPSNRTables.sh`

```

1 #!/bin/sh
2 #windows
3 source config.sh

```

```

4 TEST="$TESTS/StructuresComparison"
5
6 TS_STREAMS="$TEST/ts"
7 PSNR_TABLES="$TEST/psnr"
8
9 echo "Generating PSNR tables for comparison"
10 for BITRATE in `ls $TS_STREAMS` ; do
11     for i in `find $TS_STREAMS/$BITRATE/* -type d` ; do
12         DEST_DIR="$PSNR_TABLES/${i#$TS_STREAMS}"
13         mkdir -p $DEST_DIR
14         for j in `ls $i` ; do
15             FILE_REF="`echo $j | sed `s/[0-9]\+_CgrLevels_SINC_FILTER_SVTdec05_clipped.${
16                 BITRATE}`-[_]\+..ts//g`"
17             DEST_FILE="$DEST_DIR/${j%.ts}.txt"
18             ORIGINAL_FILE="`find $VQEG_YUV_CLIPPED_SEQS -name `*`$FILE_REF*`"
19             TS_FILE="$i/$j"
20             if [ ! -e $DEST_FILE ] ; then
21                 echo "Generating PSNR table:"
22                 echo "$DEST_FILE"
23                 if echo "$j" | grep -q "720p" ; then
24                     WIDTH="1280"; HEIGHT="720"
25                 elif echo "$j" | grep -q "1080p" ; then
26                     WIDTH="1920"; HEIGHT="1080"
27                 fi
28                 TS_FILE="`cygpath -m $TS_FILE`";
29                 ORIGINAL_FILE="`cygpath -m $ORIGINAL_FILE`"
30                 TMP_FILE="`cygpath -m $TMP/tmp.yuv`"
31                 DEST_FILE="`cygpath -m $DEST_FILE`"
32                 mencoder $TS_FILE -ovc raw -vf format=i420 -of rawvideo -o $TMP_FILE
33                 $JSVM_PATH/PSNRStatic $WIDTH $HEIGHT $ORIGINAL_FILE $TMP_FILE >
34                 $DEST_FILE
35                 rm -f $TMP/tmp.yuv
36                 echo ""
37             fi
38         done
39     done
40 done

```

B.5 Comparação Entre Estruturas HBF em Camada Simples

B.5.1 Configuração do JSVM para Camada Simples em 720p

Estes arquivos servem de base para a geração dos arquivos de configuração utilizados durante a codificação das seqüências em 720p com os roteiros da Sessão B.5.3.

(Linux): `<env_root>/scripts/tests/JSVMSingleLayer/720p_Main.cfg`

1	FrameRate	50	# Maximum frame rate [Hz]
2	FramesToBeEncoded	49	# Number of frames (at input frame rate)
3	GOPSize	4	# GOP Size (at maximum frame rate)
4	IntraPeriod	24	# Intra Period

```

5 NumberReferenceFrames 4      # Number of reference pictures
6 BaseLayerMode         1      # Base layer mode (0,1: AVC compatible,
7                          #                          2: AVC w subseq SEI)
8 SearchMode            4      # Search mode (0:BlockSearch, 4:FastSearch)
9 SearchFuncFullPel     3      # Search function full pel
10                      # (0:SAD, 1:SSE, 2:HADAMARD, 3:SAD-YUV)
11 SearchFuncSubPel     2      # Search function sub pel
12                      # (0:SAD, 1:SSE, 2:HADAMARD)
13 SearchRange           32     # Search range (Full Pel)
14 FastBiSearch          1      # Fast bi-directional search (0: off, 1:on)
15 BiPredIter            4      # Max iterations for bi-pred search
16 IterSearchRange       8      # Search range for iterations (0: normal)
17 NumLayers             1      # Number of layers
18 LayerCfg              720p_SL.cfg # Layer configuration file
19 PreAndSuffixUnitEnable 0     # Add prefix unit (0: off, 1: on)
20 MMCOBaseEnable       0      # MMCO for base representation (0: off, 1: on)

```

(Linux): <env_root>/scripts/tests/JSVMSingleLayer/720p_SL.cfg

```

1 SourceWidth           1280    # Input frame width
2 SourceHeight          720     # Input frame height
3 FrameRateIn           50      # Input frame rate [Hz]
4 FrameRateOut          50      # Output frame rate [Hz]
5 ReconFile             /dev/null # Reconstructed file
6 ProfileIdc            100     # value of profile_idc (or 0 for auto detection)
7 Enable8x8Transform    1      # 8x8 luma trafo (0:disabled, 1:enabled)
8 SliceMode             1      # (0=off 1=fixed #mb in slice
9                          # 2=fixed #bytes in slice)
10 SliceArgument         3600    # (Arguments to modes 1 and 2 above)

```

B.5.2 Configuração do JSVM para Camada Simples em 1080p

Estes arquivos servem de base para a geração dos arquivos de configuração utilizados durante a codificação das seqüências em 1080p com os roteiros da Sessão B.5.3.

(Linux): <env_root>/scripts/tests/JSVMSingleLayer/1080p_Main.cfg

```

1 FrameRate             50      # Maximum frame rate [Hz]
2 FramesToBeEncoded     49     # Number of frames (at input frame rate)
3 GOPSize               4      # GOP Size (at maximum frame rate)
4 IntraPeriod           24     # Intra Period
5 NumberReferenceFrames 4      # Number of reference pictures
6 BaseLayerMode         1      # Base layer mode (0,1: AVC compatible,
7                          #                          2: AVC w subseq SEI)
8 SearchMode            4      # Search mode (0:BlockSearch, 4:FastSearch)
9 SearchFuncFullPel     3      # Search function full pel
10                      # (0:SAD, 1:SSE, 2:HADAMARD, 3:SAD-YUV)
11 SearchFuncSubPel     2      # Search function sub pel
12                      # (0:SAD, 1:SSE, 2:HADAMARD)
13 SearchRange           32     # Search range (Full Pel)
14 FastBiSearch          1      # Fast bi-directional search (0: off, 1:on)
15 BiPredIter            4      # Max iterations for bi-pred search
16 IterSearchRange       8      # Search range for iterations (0: normal)

```

```

17 NumLayers          1          # Number of layers
18 LayerCfg           1080p_SL.cfg # Layer configuration file
19 PreAndSuffixUnitEnable 0      # Add prefix unit (0: off, 1: on)
20 MMCOBaseEnable     0          # MMCO for base representation (0: off, 1: on)

```

(Linux): <env_root>/scripts/tests/JSVMSingleLayer/1080p_SL.cfg

```

1 SourceWidth        1920        # Input frame width
2 SourceHeight       1080        # Input frame height
3 FrameRateIn        50          # Input frame rate [Hz]
4 FrameRateOut       50          # Output frame rate [Hz]
5 ReconFile          /dev/null    # Reconstructed file
6 ProfileIdc         100         # value of profile_idc (or 0 for auto detection)
7 Enable8x8Transform 1          # 8x8 luma trafo (0:disabled, 1:enabled)
8 SliceMode          1          # (0=off 1=fixed #mb in slice
9                   # 2=fixed #bytes in slice)
10 SliceArgument      3600        # (Arguments to modes 1 and 2 above)

```

B.5.3 Roteiros para Codificação das Camadas Simples com o JSVM

Estes roteiros geram, a partir dos arquivos de configuração (B.5.1 e B.5.2), arquivos temporários para codificação das seqüências de vídeo em camada simples com o JSVM, com variações dos parâmetros de intervalo de quadros (24 e 48), de tamanho de GOP (4 e 8) e de coeficiente de referência (24, 28, 32, 36 e 40 para 720p; 28, 32, 36 e 40 para 1080p). Os fluxos e os relatórios de codificação resultantes são armazenados em <env_root>/tests/JSVMSingleLayer/, organizados por diretórios criados pela combinação de intervalo de quadros e tamanho de GOP.

(Linux): <env_root>/scripts/tests/JSVMSingleLayer/encode.sh

```

1 #!/bin/bash
2 CPUS_LOCK_FILE=cpus_lock.file
3
4 function unlock_cpu() {
5     if [ -f "$CPUS_LOCK_FILE" ] ; then
6         CPUS_IN_USE='cat $CPUS_LOCK_FILE'
7         CPUS_IN_USE=$((--CPUS_IN_USE))
8         echo "CPU unlocked ($CPUS_IN_USE in use)"
9         echo "$CPUS_IN_USE" > ${CPUS_LOCK_FILE}.tmp
10        mv ${CPUS_LOCK_FILE}.tmp ${CPUS_LOCK_FILE}
11    fi
12 }
13
14 /usr/bin/time -o "$8" "$JSVM_PATH/H264AVCEncoderLibTestStatic" -pf "$1" -bf "$2" -gop "$3"
15    -iper "$4" -org 0 "$5" -lqp 0 "$6" > "$7"
16 unlock_cpu

```

(Linux): <env_root>/scripts/tests/JSVMSingleLayer/encodeSingleLayers.sh

```

1  #!/bin/bash
2  source "$PWD/config.sh"
3  TEST="$TESTS/JSVMSingleLayer"
4  TEST_SCRIPTS="$PWD/$(dirname $0)"
5  CPUS='cat /proc/cpuinfo | awk '/cpu cores/ { if(x++ == 0) print $4 }'
6  CPUS_LOCK_FILE=cpus_lock_file
7
8  function lock_cpu() {
9      CPUS_IN_USE=0
10     if [ -f "$CPUS_LOCK_FILE" ] ; then
11         CPUS_IN_USE='cat $CPUS_LOCK_FILE'
12     fi
13     if [ $CPUS_IN_USE -lt $CPUS ] ; then
14         CPUS_IN_USE=$((++CPUS_IN_USE))
15         echo "$CPUS_IN_USE" > ${CPUS_LOCK_FILE}.tmp
16         mv ${CPUS_LOCK_FILE}.tmp ${CPUS_LOCK_FILE}
17         echo "CPU locked ($CPUS_IN_USE in use)"
18     else
19         return 1
20     fi
21 }
22
23 function encode() {
24     REPORT_FILES="$TEST}/I${1}_GOP${2}/doc/jsvm"
25     ES_FILES="$TEST}/I${1}_GOP${2}/264"
26     TIME_FILES="$TEST}/I${1}_GOP${2}/time"
27
28     for i in `find $VQEG_YUV_CLIPPED_SEQS/* -type d` ; do
29         REPORT_DIR="$REPORT_FILES/${i#$VQEG_YUV_CLIPPED_SEQS}"
30         ES_DIR="$ES_FILES/${i#$VQEG_YUV_CLIPPED_SEQS}"
31         TIME_DIR="$TIME_FILES/${i#$VQEG_YUV_CLIPPED_SEQS}"
32         for j in `ls $i` ; do
33             SRC_FILE="$i/$j"
34             RESOLUTION='echo $i | sed -e 's,.*\/(.*)\,1,' -e 's\/\([0-9]\+[ip]\)\.*/1/'
35             if [ "$RESOLUTION" = "720p" ] ; then QP=${QP}_720p[@]
36             elif [ "$RESOLUTION" = "1080p" ] ; then QP=${QP}_1080p[@]
37             else continue
38             fi
39             for k in ${QP[@]} ; do
40                 ES_FILE="$ES_DIR/${j%.yuv}QP${k}_264"
41                 if [ -e $ES_FILE ] ; then
42                     echo "File already generated: $ES_FILE"
43                     continue
44                 fi
45                 echo "Encoding file: $ES_FILE"
46                 mkdir -p "$REPORT_DIR" "$ES_DIR" "$TIME_DIR"
47                 CFG_FILE="$TEST_SCRIPTS}/${RESOLUTION}_Main.cfg"
48                 REPORT_FILE="$REPORT_DIR/${j%.yuv}QP${k}_txt"
49                 TIME_FILE="$TIME_DIR/${j%.yuv}QP${k}_txt"
50                 while true ; do
51                     lock_cpu
52                     if [ "$?" == 0 ] ; then break ; else sleep 3 ; fi
53                 done
54                 JSVM_PATH=$JSVM_PATH ./encode.sh "$CFG_FILE" "$ES_FILE" "$2" "$1" "$SRC_FILE" "$k" "$REPORT_FILE" "$TIME_FILE" &
55                 echo ""
56             done

```

```

57     done
58     done
59 }
60
61 QP_720p=(24 28 32 36 40)
62 QP_1080p=(28 32 36 40)
63 IFR=(24 48)
64 GOP=(4 8)
65
66 pushd "$TEST_SCRIPTS"
67 echo "Generating VBR streams"
68 rm -f $CPUS_LOCK_FILE
69 for g in ${IFR[@]} ; do
70     for f in ${GOP[@]} ; do
71         encode "$g" "$f"
72     done
73 done
74 popd

```

B.5.4 Configuração do AVISynth para Medição SSIM

Este arquivo serve de base para a geração dos arquivos utilizados durante a decodificação das seqüências geradas pelo JSVM com o roteiro da Sessão B.5.5 para análise SSIM.

(Windows): <env_root>/scripts/tests/JSVMSingleLayer/ssim.avs

```

1 Original=RawSource("ORIG", WIDTH, HEIGHT, "I420").trim(0,48)
2 Reconst=RawSource("RECONST", WIDTH, HEIGHT, "I420")
3 ssim(Reconst,Original,"RESULTS","TMP/averageSSIM.txt")

```

B.5.5 Roteiro para Geração das Medidas SSIM das Codificações Simples com o JSVM

Este roteiro realiza o levantamento das medidas SSIM das seqüências codificadas com o JSVM em camada simples, armazenando os resultados na mesma estrutura de diretórios criada para os fluxos codificados e relatórios, <env_root>/tests/JSVMSingleLayer/. Durante a execução do roteiro, o arquivo da Sessão B.5.4 é utilizado para gerar arquivos temporários que são executados com o Windows Media Player TM, que executa o AVISynth por baixo.

(Windows): <env_root>/scripts/tests/JSVMSingleLayer/genSSIMDocs.sh

```

1 #!/bin/bash
2 source "config.sh"
3 TEST="$TESTS/JSVMSingleLayer"
4 TEST_SCRIPTS="$PWD/`dirname $0`"
5

```

```

6  echo "Generating SSIM Docs"
7  for STRUCT in `ls $TEST` ; do
8      ENC_FILES="$TEST/$STRUCT/264"
9      AVISYNTH_FILES="$TEST/$STRUCT/doc/avisynth"
10     for i in `find $ENC_FILES/* -type d` ; do
11         RESOLUTION=`echo $i | sed -e 's,*/\(.*\),\1,' -e 's/\([0-9]\+[ip]\)\.*/\1/'`
12         if [ "$RESOLUTION" = "720p" ] ; then
13             WIDTH="1280"
14             HEIGHT="720"
15         elif [ "$RESOLUTION" = "1080p" ] ; then
16             WIDTH="1920"
17             HEIGHT="1080"
18         else continue
19         fi
20         AVISYNTH_DIR="$AVISYNTH_FILES/${i#$ENC_FILES}"
21         mkdir -p "$AVISYNTH_DIR"
22         for j in `ls $i` ; do
23             FILE_RADICAL=`echo $j | sed -e 's/QP[0-9]\+_/ /' -e 's/\..264//`
24             AVISYNTH_FILE="$AVISYNTH_DIR/${j%.264}.csv"
25             if [ ! -e "$AVISYNTH_FILE" ] ; then
26                 echo "Generating SSIM doc:"
27                 echo "$AVISYNTH_FILE"
28                 ES_FILE=`cygpath -m `"$i/$j`""
29                 "$JSVM_PATH/H264AVCDecoderLibTestStatic" "$ES_FILE" "`cygpath -w `"$i/$j`""
30                 ORIGINAL=`cygpath -m $VQEG_YUV_CLIPPED_SEQS/${i#$ENC_FILES}/${FILE_RADICAL}.yuv`
31                 RECONST=`cygpath -m $TMP/rec.yuv`
32                 RESULTS=`cygpath -m $AVISYNTH_FILE`
33                 cat "$TEST_SCRIPTS/ssim.avs" | sed -e "s,ORIG,$ORIGINAL," -e "s,RECONST,$RECONST," -e "s,WIDTH,$WIDTH," -e "s,HEIGHT,$HEIGHT," -e "s,RESULTS,$RESULTS," -e "s,TMP,'$TMP'," > "$TMP/ssim.avs"
34                 "$MPLAYER6" /play /close "`cygpath -m $TMP/ssim.avs`"
35             else
36                 echo "File $AVISYNTH_FILE already generated"
37             fi
38         done
39     done
40 done
41 rm -rf "$TMP/rec.yuv" "$TMP/averageSSIM.txt" "$TMP/ssim.avs"

```

B.6 Comparação Entre Escalabilidade e *Simulcast*

B.6.1 Configuração do JSVM para Codificação Escalável

Estes arquivos servem de base para a geração dos arquivos de configuração utilizados durante a codificação das seqüências escaláveis de 720p para 1080p com o roteiro da Sessão B.6.2.

(Linux): <env_root>/scripts/tests/JSVMScalableLayers/720p_1080p_Main.cfg

```
1 # JSVM Main Configuration File
```

```

2
3  OutputFile          OUT_FILE # Bitstream file
4  FrameRate          50      # Maximum frame rate [Hz]
5  FramesToBeEncoded  49      # Number of frames (at input frame rate)
6  GOPSize            8       # GOP Size (at maximum frame rate)
7  IntraPeriod        24      # Intra Period
8  NumberReferenceFrames 4      # Number of reference pictures
9  BaseLayerMode      1       # Base layer mode (0,1: AVC compatible,
10                          #                          2: AVC w subseq SEI)
11  SearchMode         4       # Search mode (0:BlockSearch, 4:FastSearch)
12  SearchRange        32      # Search range (Full Pel)
13  NumLayers          2       # Number of layers
14  LayerCfg           720p_L0.cfg # Layer configuration file
15  LayerCfg           1080p_L1.cfg # Layer configuration file

```

(Linux): <env_root>/scripts/tests/JSVMScalableLayers/720p_L0.cfg

```

1 # JSVM Layer Configuration File
2 SourceWidth        1280    # Input frame width
3 SourceHeight       720     # Input frame height
4 FrameRateIn        50     # Input frame rate [Hz]
5 FrameRateOut       50     # Output frame rate [Hz]
6 InputFile          IN_FILE # Input file
7 ReconFile          /dev/null # Reconstructed file

```

(Linux): <env_root>/scripts/tests/JSVMScalableLayers/1080p_L1.cfg

```

1 # JSVM Layer Configuration File
2 SourceWidth        1920    # Input frame width
3 SourceHeight       1080    # Input frame height
4 FrameRateIn        50     # Input frame rate [Hz]
5 FrameRateOut       50     # Output frame rate [Hz]
6 InputFile          IN_FILE # Input file
7 ReconFile          /dev/null # Reconstructed file
8 InterLayerPred     2       # Inter-layer Pred. (0: no, 1: yes, 2: adap.)
9 UseESS             1       # ESS mode
10 ESSCropWidth       1920    # base layer upsampled frame width
11 ESSCropHeight      1080    # base layer upsampled frame height
12 ESSOriginX         0       # base layer upsampled frame x-pos
13 ESSOriginY         0       # base layer upsampled frame y-pos

```

B.6.2 Roteiro para Codificação Escalável com o JSVM

Estes roteiros geram, a partir dos arquivos de configuração (B.6.1), arquivos temporários para codificação das seqüências de vídeo com escalabilidade de um camada 720p para uma camada 1080p com o JSVM. Para as codificações são utilizados intervalo de quadros 24, tamanho de GOP 8 e coeficientes de referência de acordo com a Tabela 5.8. Os fluxos e os relatórios de codificação resultantes são armazenados em <env_root>/tests/JSVMScalableLayers/.

(Linux): <env_root>/scripts/tests/JSVMScalableLayers/encodeScalableLayers.sh

```

1  #!/bin/bash
2  source "$PWD/config.sh"
3  TEST="$TESTS/JSVMScalableLayers"
4  TEST_SCRIPTS="$PWD/`dirname $0`"
5
6  DEST_DIR="$TEST/264/svc"
7  DOC_DIR="$TEST/doc/svc"
8  TIME_DIR="$TEST/time/svc"
9  mkdir -p "$DEST_DIR"; mkdir -p "$DOC_DIR"; mkdir -p "$TIME_DIR"
10
11  function encode() {
12      pushd "$TEST_SCRIPTS"
13      SRC_FILE_720p=`find "$VQEG_YUV_CLIPPED_SEQS" -type f -name \ "$1*720p*\`"
14      SRC_FILE_1080p=`find "$VQEG_YUV_CLIPPED_SEQS" -type f -name \ "$1*1080p*\`"
15      DEST_FILE="$DEST_DIR/${1}_720p_Qp${2}_1080p_Qp${3}.264"
16      DOC_FILE="$DOC_DIR/${1}_720p_Qp${2}_1080p_Qp${3}.txt"
17      TIME_FILE="$TIME_DIR/${1}_720p_Qp${2}_1080p_Qp${3}.txt"
18      /usr/bin/time -o "$TIME_FILE" "$JSVM_PATH/H264AVCEncoderLibTestStatic" -pf "720
19          p_1080p_Main.cfg" -bf "$DEST_FILE" -org 0 "$SRC_FILE_720p" -lqp 0 "$2" -org 1 "
20          $SRC_FILE_1080p" -lqp 1 "$3" > "$DOC_FILE"
21      popd
22  }
23
24  encode CrowdRun 36 38
25  encode ParkJoy 37 39
26  encode DucksTakeOff 36 36
27  encode InToTree 33 32
28  encode OldTown 34 32

```