



Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Departamento de Engenharia Elétrica
Programa de Pós-Graduação em Engenharia Elétrica

Dissertação de Mestrado

**Avaliação de Técnicas de Redução da
Resolução Espacial de Vídeos para
Dispositivos Móveis**

Carlos Danilo M. Regis

Campina Grande – PB
Março - 2009

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Departamento de Engenharia Elétrica
Programa de Pós-Graduação em Engenharia Elétrica

Avaliação de Técnicas de Redução da Resolução Espacial de Vídeos para Dispositivos Móveis

Carlos Danilo M. Regis

Dissertação de Mestrado submetida à Coordenação do Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da Universidade Federal de Campina Grande como requisito necessário para obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Comunicações.

Marcelo Sampaio de Alencar
Orientador

Mylène Christine Queiroz de Farias
Orientadora

Campina Grande – PB, Paraíba, Brasil
©Carlos Danilo M. Regis

*A Deus!
A meus pais, Roberto e Carla, meus irmãos, Caio e Tiago e
a minha noiva Gabriella.*

Agradecimentos

Agradeço primeiramente a Deus pelo seu amor por mim, por toda a ajuda nos momentos mais difíceis, e por nunca me deixar faltar nada.

Aos meus pais, Roberto e Carla e aos meus irmãos Caio e Tiago, por serem sempre presente em minha vida, com amor, cuidado, ensinamentos, broncas (que me fizeram crescer) e confiança durante todo o tempo.

À minha amada noiva Gabriella, que me motivou a procurar a pesquisa e acreditou que eu poderia chegar até esse momento, ainda estando na graduação. Sou grato pelo incentivo e apoio durante toda a execução desse trabalho e pelo amor, carinho e compreensão demonstrados em todos os períodos desde a graduação.

Aos meus tios Ricardo e Cláudia que me acolheram como um filho, me dando todas as condições necessárias para chegar até aqui.

Aos meus orientadores Marcelo Sampaio de Alencar e Mylène Christine Queiroz de Farias, pela orientação deste trabalho, estímulo e dedicação sempre prestados, que muito me enriqueceram, fortalecendo o meu desenvolvimento profissional.

Aos meus grandes amigos, Rogério, Alberto e Jerônimo, que sempre estiveram prontos para me ouvir e compartilhar os momentos da minha vida.

Aos meus amigos do laboratório, Paulo, Erik, Késia, Portela e Fausy, pela amizade e apoio durante o trabalho no laboratório.

Aos amigos Raíssa e Daniel que colaboraram muito com o desenvolvimento da dissertação.

A CAPES pelo aporte financeiro, sem o qual seria impossível a realização deste trabalho.

Ao Iecom, que forneceu toda a infraestrutura necessária para a realização deste trabalho.

Resumo

Esta dissertação aborda a avaliação da qualidade de sinais de vídeo, baseada na avaliação objetiva e subjetiva, para vídeos utilizados em televisão digital para dispositivos móveis. Essa necessidade aparece com o surgimento da televisão digital, que permite a manipulação do vídeo para diferentes dispositivos, incluindo dispositivos móveis. No Brasil o padrão adotado define uma banda do espectro para a transmissão da televisão para esses dispositivos, limitada em 433 kHz. Em função desse limite faz-se necessário o ajuste de alguns parâmetros, como taxa de *bit* e a resolução espacial e temporal. Para o ajuste desses parâmetros usam-se transcodificadores de vídeos, que são sistemas capazes de converter uma sequência de vídeo em uma outra com diferentes parâmetros. Neste trabalho é feita a redução espacial dos vídeos de CIF (352×288) para QCIF (176×144). Para a redução espacial de vídeo o transcodificador pode ser implementado usando vetores de movimentos, DCT ou interpolador fracional. O interpolador fracional foi escolhido por ter baixo custo computacional. Ele usa os filtros da média, mediana, moda, sigma e média ponderada, com janelas de 2×2 , 3×3 e 4×4 . Para a avaliação da qualidade dos vídeos transcodificados foram usados os métodos objetivos PSNR e SSIM e o método subjetivo PC. Para a avaliação subjetiva foi utilizado o codificador H.264 para a redução da taxa de *bit* e da resolução temporal do vídeo, que foi apresentado em um aparelho celular.

Palavras-chave: TV Digital, TV Móvel, Vídeo digital, Transcodificação, Filtros, Processamento de Sinais de Vídeo.

Abstract

This dissertation evaluates the quality of video signals. It is based on subjective and objective evaluation of videos used in digital television for mobile devices. The emergence of digital television implies the need to manipulate videos, in order to adapt them to various devices. Brazil adopted a standard that allocates one band of spectrum for the transmission of television signal for mobile television devices, with a bandwidth of 433 kHz. Because of the bandwidth restriction it is necessary to adjust some parameters such as bit rate and spatial and temporal resolution. Transcoded videos are obtained from the conversion of a video sequence in another with different parameters. This dissertation studies the spatial compression of videos from CIF (352×288) to QCIF (176×144). To reduce the video a transcoder can be implemented using motion vectors, DCT or fractional interpolator. The fractional interpolator was chosen to be implemented because it has low computational complexity. It employs the average, median, mode and weighted average filters, with windows of 2×2 , 3×3 and 4×4 . The objective PSNR and SSIM and subjective PC methods, were used to evaluate the quality of the transcoded videos. The subjective evaluation used the H.264 encoder with reduced bit rate and temporal resolution of the video, implemented using a cellular device.

Keywords: Digital TV, Mobile TV, Digital video, Transcoding, Filters.

Sumário

1	Introdução	1
1.1	Objetivos	2
1.2	Motivação	3
1.3	Contribuições	5
1.4	Estrutura do Texto	5
2	Televisão Móvel	6
2.1	Transmissão da Televisão Móvel	7
2.1.1	Televisão Móvel em Redes Celulares	8
2.1.2	Televisão Móvel em Redes de Televisão Digital	8
2.1.3	Parâmetros de Avaliação dos Padrões	9
2.2	Padrões de Televisão Móvel	9
2.3	One-Seg	10
2.3.1	Parâmetros de Transmissão	11
2.4	Televisão Móvel no Brasil	12
2.4.1	Receptor de Televisão Móvel	12
2.4.2	Codificação de Vídeo	13
2.4.3	Resoluções de Luminância	14
2.4.4	Mercado Atual Brasileiro	15
2.5	Aceitabilidade e Avaliação da Qualidade Visual	16
3	Sistema de Transcodificação	17
3.1	Tipos de Transcodificação	18
3.1.1	Transcodificação Homogênea	19
3.1.2	Transcodificação Heterogênea	20
3.2	Filtros no Domínio Espacial	20
3.2.1	Princípio de Funcionamento	21
3.2.2	Definição de Vizinhança	22
3.2.3	Tipos de Filtros	22
3.2.4	Filtros Passa Baixas	23

3.3	Sistema Proposto	28
4	Métodos de Avaliação	31
4.1	Medidas Objetivas	31
4.2	Medidas Subjetivas	35
4.3	Métricas Estatísticas	37
4.3.1	Correlação Entre as Medidas	38
5	Resultados Experimentais	39
5.1	Análise dos Vídeos em Diferentes Dispositivos	40
5.2	Avaliação da Taxa de <i>Bit</i> de Vídeo	44
5.3	Medidas Objetivas dos Vídeos Transcodificados	46
5.3.1	MSE e PSNR	46
5.3.2	SSIM	50
5.3.3	Tempo de Processamento	50
5.3.4	Erro Provocado pela Codificação	52
5.4	Testes Subjetivos dos Vídeos Transcodificados	54
5.4.1	Avaliação do Primeiro Grupo de Vídeos	55
5.4.2	Avaliação do Segundo Grupo de Vídeos	57
5.4.3	Avaliação Geral dos Dados Subjetivos	58
6	Conclusões	61
A	Medidas Subjetivas	63
A.1	Medidas Subjetivas para Aplicações em Televisão	63
A.1.1	Método DSIS	64
A.1.2	Método DSCQS	65
A.1.3	Método SS	66
A.1.4	Métodos de Comparação de Estímulos	66
A.1.5	Método SSCQE	67
A.1.6	Método SDSCE	67
A.2	Medidas Subjetivas para Aplicações Multimídia	68
A.2.1	Método ACR	68
A.2.2	Método DCR	69
B	O Codificador H.264	70
B.1	O Filtro de Anti-blocagem	72
B.2	CABAC/CAVLC	72
B.3	Tamanho de Bloco Variável / Partições de Macroblocos	73
B.4	Compensação de movimento com múltiplos quadros de referência	73
B.5	Otimização da Taxa de Distorção	73

C	Questionário - TV Móvel	74
C.1	Expectativas dos Usuários sobre a TV móvel	76
D	Artigos Publicados	77

Lista de Figuras

1.1	Sistema básico de TV.	2
2.1	Transmissão para TV móvel por radiodifusão e <i>unicast</i>	7
2.2	Serviço ISDB-T.	11
2.3	Configuração básica do receptor.	13
2.4	Aparelho V820L, para transmissão de TV Móvel (SAMSUNG, 2008).	16
3.1	Modelo de um transcodificador de vídeo.	17
3.2	Interpolador fracional com taxa de L/S.	18
3.3	Diagrama dos possíveis tipos de transcodificação.	19
3.4	Máscara de deslocamento.	21
3.5	Representação da vizinhança do pixel central com valor x_s	23
3.6	Transcodificação espacial com janela 2×2 . (a) Quadro CIF 352×288 , (b) Quadro QCIF 176×144	29
3.7	Técnica da média de <i>pixels</i> com janela 3×3 . (a) Quadro CIF 352×288 , (b) Quadro QCIF 176×144	30
3.8	Processo de comparação do desempenho de vídeos transcodificados.	30
4.1	Diagrama simplificado para medição da similaridade estrutural entre duas imagens.	33
4.2	Apresentação dos estímulos no método PC.	37
5.1	Vídeos Paris, Mobile e Tempete.	40
5.2	Gráfico da PSNR média versus dimensão da janela usando o filtro da média.	41
5.3	Compressão versus dimensão da janela usando o filtro da média.	41
5.4	Valor médio dos valores dados para cada vídeo usando o N800.	43
5.5	Valor médio dos valores (MOS) dados para cada vídeo usando o N95, com taxa de <i>bit</i> de 243 <i>kbit/s</i>	44
5.6	Vídeos Walk, Akiyo e Coastguard	44
5.7	MOS obtidos para cada uma das sequencias de teste.	46
5.8	Valor do MOS para os vídeos com taxa de 318 <i>kbit/s</i>	47

5.9	Vídeo Mobile, News e Foreman	47
5.10	Gráfico da MSE para os vídeos transcodificados.	49
5.11	Gráfico da PSNR para os vídeos transcodificados.	49
5.12	Gráfico do SSIM para os vídeos transcodificados.	51
5.13	Gráfico do PSNR para dos vídeos codificados após a transcodificação.	53
5.14	Gráfico do SSIM para os vídeos codificados após a transcodificação.	55
5.15	Valores do MOS para o primeiro grupo de vídeos transcodificados.	56
5.16	Valores do MOS para o segundo grupo de vídeos transcodificados.	58
5.17	Valores do MOS para o vídeo Foreman.	59
5.18	Valores do MOS para o vídeo Mobile.	59
5.19	Valores do MOS para o vídeo News.	60
A.1	Tempos de observação e julgamento, método DSIS.	64
A.2	Classificação da qualidade da imagem pelo método DSCQS.	65
A.3	Apresentação dos estímulos no método DCR.	69
B.1	Estrutura completa de um codificador H.264/AVC.	71
B.2	Estrutura completa de um decodificador H.264/AVC.	71
B.3	Comparação entre um quadro filtrado e um quadro sem filtragem.	72
B.4	Referência múltipla.	73

Lista de Tabelas

2.1	Comparação das tecnologias de Televisão Móvel baseadas em transmissão por radiodifusão	10
2.2	Parâmetros de modulação: QPSK, intervalo de guarda de 1/8, mode3, corretor de erro convolucional de 1/2.	12
2.3	Restrições aos parâmetros de codificação.	14
2.4	Limitações impostas por cada nível.	14
4.1	PVD para imagens em movimento	35
5.1	MOS, desvio padrão e coeficiente de variação dos testes subjetivos para o N800.	42
5.2	MOS, desvio padrão e coeficiente de variação dos testes subjetivos para o N95.	43
5.3	MOS, desvio padrão e o coeficiente de variação dos vídeo originais com diferentes taxas.	45
5.4	Valores do MOS, desvio padrão e coeficiente de variação dos MOS obtidos dos testes subjetivos utilizando o N95 e taxa de <i>bit</i> de 318 <i>kbit/s</i>	46
5.5	Resultados do MSE.	48
5.6	Resultados do PSNR.	48
5.7	Resultados do método SSIM.	50
5.8	Tempo de processamento de um vídeo	51
5.9	Resultados do PSNR dos vídeos codificados após a transcodificação	52
5.10	Resultados dos vídeos codificados após a transcodificação utilizando o método SSIM.	54
5.11	Valores do MOS, desvio padrão e CV dos dados coletados do primeiro grupo de testes subjetivos para os vídeos transcodificados.	55
5.12	Valores do MOS, desvio padrão e CV dos dados coletados do segundo grupo de testes subjetivos para os vídeos transcodificados.	57
A.1	conceito para avaliação DSIS.	65
A.2	Conceito para avaliação no método “Comparação de estímulos”.	66

Lista de Siglas

2G	<i>Second Generation of Mobile Communications</i>	Segunda Geração de Comunicações Móveis
2,5G	<i>Second and Half Generation of Mobile Communications</i>	Geração 2,5 de Comunicações Móveis
3G	<i>Third Generation of Mobile Communications</i>	Terceira Geração de Comunicações Móveis
AAC	<i>Advanced Audio Coding</i>	Codificação de Áudio Avançado
ACR	<i>Absolute Category Rating</i>	Ordenamento por Categoria Absoluta
ASO	<i>Arbitrary Slice Ordering</i>	Ordenamento Arbitrário de Fatias
BSAC	<i>Bit Sliced Arithmetic Coding</i>	Codificação Aritmética com Fatias de bits
CBR	<i>Constant Bit-Rate</i>	Taxa de bit Constante
CIF	<i>Common Intermediate Format</i>	Formato Intermediário Comum
CV	<i>Coefficients of Variation</i>	Coefficiente de Variação
COFDM	<i>Coded Orthogonal Frequency-Division Multiplexing</i>	Multiplexação Ortogonal por Divisão em Frequência Codificado
DAB	<i>Digital Audio Broadcasting</i>	Transmissão de Áudio Digital
DCR	<i>Degradation Category Rating</i>	Ordenamento por Categoria de Degradação
DCT	<i>Discrete Cosine Transform</i>	Transformada Discreta do Cosseno
DMB	<i>Digital Multimedia Broadcast</i>	Transmissão Digital de Multimídia
DTTB	<i>Digital Terrestrial Television Broadcasting</i>	Transmissão de Televisão Digital Terrestre
DQPSK	<i>Differential Quadrature Phase Shift Keying</i>	Chaveamento por Deslocamento de Fase em Quadratura Diferencial
DVB-H	<i>Digital Video Broadcasting - Handheld</i>	Transmissão de Vídeo Digital para Receptores Portáteis
DVB-T	<i>Digital Video Broadcasting - Terrestrial</i>	Transmissão de Vídeo Digital Terrestre

DVB-SH	<i>Digital Video Broadcasting - Satellite services to Handhelds</i>	Transmissão de Vídeo Digital para Receptores Portáteis Usando Satélites
DSCQS	<i>Double Stimulus Continuous Quality Scale</i>	Escala de Qualidade de Duplo Estímulo Contínuo
DSIS	<i>Double Stimulus Impairment Scale</i>	Escala de Restrição com Duplo Estímulo
EDGE	<i>Enhanced Data Rate for GSM Evolution</i>	Taxas Elevadas de Transmissão de Dados para a Evolução do GSM
ETSI	<i>European Telecommunications Standard Institute</i>	Instituto Europeu de Padrões de Telecomunicações
FDM	<i>Frequency Division Multiplexing</i>	Multiplexação por Divisão em Frequência
FM	<i>Frequency Modulation</i>	Modulação em Frequência
FMO	<i>Flexible Macroblock Ordering</i>	Ordenação Flexível de Grandes Blocos
HDTV	<i>High-Definition Television</i>	Televisão de Alta Definição
HVS	<i>Human Visual System</i>	Sistema Visual Humano
INdT	<i>NOKIA Institute of Tech</i>	Instituto NOKIA de Tecnologia
ISDB-T	<i>Integrated Services Digital Broadcasting Terrestrial</i>	Serviço Integrado de Transmissão Digital Terrestre
ISDB-Tb	<i>Integrated Services Digital Broadcasting Terrestrial Built-in</i>	Serviço Integrado de Transmissão Digital Terrestre Customizado
ISDTV	<i>International System for Digital Television</i>	Sistema Internacional para Televisão Digital
ITU-R	<i>International Telecommunication Union</i>	União Internacional de Telecomunicações
ITU-T	<i>International Telecommunication Union</i>	União Internacional de Telecomunicações
LAN	<i>Local Area Network</i>	Rede em Área Local
MBMS	<i>Multimedia Broadcast Multicast Service</i>	Serviço de Transmissão <i>Multicast</i> Multimídia
MOS	<i>Mean Opinion Score</i>	Opinião Média da Avaliação
MPEG-2	<i>Moving Picture Experts Group - 2</i>	Grupo Especializado em Imagens Móveis
MPEG-TS	<i>Moving Picture Experts Group - Transport Stream</i>	Grupo Especializado em Imagens Móveis - Fluxo de Transporte
MSE	<i>Mean Squared Error</i>	Erro Médio Quadrático
OFDM	<i>Orthogonal Frequency-Division Multiplexing</i>	Multiplexação Ortogonal por Divisão em Frequência

OLED	<i>Organic Light-Emitting Diode</i>	diodo orgânico emissor de luz
One-Seg	<i>One Segments</i>	Um Segmento
PC	<i>Pair Comparison</i>	Comparação em Pares
PCM	<i>Pulse-Code Modulation</i>	Modulação por Código de Pulsos
PDA	<i>Personal Digital Assistants</i>	Assistente Pessoal Digital
PSNR	<i>Peak Signal-to-Noise Ratio</i>	Razão Sinal de Pico Ruído
PVD	<i>Preferred Viewing Distance</i>	Distância de Exibição preferencial
QAM	<i>Quadrature Amplitude Modulation</i>	Modulação de Amplitude em Quadratura
QCIF	<i>Quarter Common Intermediate Format</i>	Um Quarto do Formato Intermediário Comum
QPSK	<i>Quadrature Phase Shift Keying</i>	Chaveamento por Deslocamento de Fase em Quadratura
QoE	<i>Quality-of-experience</i>	Qualidade do Experimento
QVGA	<i>Quarter Video Graphics Array</i>	Arranjo Gráfico de um Quarto de Vídeo
RS	<i>Redundant Slices</i>	Fatias Redundantes
SDTV	<i>Stand-Definition Television</i>	Televisão de Definição padrão
SNR	<i>Signal-to-Noise Ratio</i>	Razão Sinal Ruído
SQVGA	<i>S Quarter Video Graphics Array</i>	Arranjo Gráfico de Sub um Quarto de Vídeo
SS	<i>Single Stimulus</i>	Estímulo Simples
SSCQE	<i>Single Stimulus Continuous Quality Evaluation</i>	Avaliação de Qualidade de Estímulo Simples Contínuo
SSIM	<i>Structural Similarity</i>	Similaridade Estrutural
UHF	<i>Ultra High Frequency</i>	Frequência Ultra Alta
UMTS	<i>Universal Mobile Telecommunication System</i>	Sistema de Telecomunicações Móveis Universal
VBR	<i>Variable Bit-Rate</i>	Taxa de <i>bit</i> Variável
VHF	<i>Very High Frequency</i>	Frequência Muito Alta
WiBro	<i>Wireless Broadband</i>	Rede sem Fio de Banda Larga
WIMAX	<i>Worldwide Interoperability for Microwave Access</i>	Interoperabilidade Mundial para Acesso por Microondas
WSNR	<i>Weighted Signal to Noise Ratio</i>	Razão entre Sinal e Ruído Ponderada

Lista de Símbolos

L	Superamostrador
S	Subamostrador
x	Variável no eixo horizontal
y	Variável no eixo vertical
j	Variável no eixo horizontal da vizinhança
l	Variável no eixo vertical da vizinhança
a	Peso dado ao elemento
b	Peso dado ao elemento
$g(x,y)$	Quadro filtrado
$f(x,y)$	Quadro original
M	Número de linhas do quadro
N	Número de colunas do quadro
W	Número de <i>pixels</i> da janela
w	Número de linhas e colunas da janela
$\alpha(x,y)$	Peso do <i>pixel</i>
K	Número dos níveis de cinza de uma janela
p	Valor do pixel
z	Valor do pixel arranjado em ordem crescente
z_m	Valor do pixel de maior frequência da janela
σ	Desvio padrão
μ_f	Média dos valores de $f(x,y)$
T	Faixa de intensidade
$f_T(x,y)$	Quadro limitado pela faixa de intensidade
P	Número total de <i>pixels</i>
k	Número total de quadros
$f(x,y,k)$	Vídeo original
$h(x,y,k)$	Vídeo transcodificado
σ_f^2	Desvio padrão de $f(x,y,k)$
σ_h^2	Desvio padrão de $h(x,y,k)$

μ_h	Média dos valores de $h(x, y, k)$
σ_{fh}	Covariância entre $f(x, y, k)$ e $h(x, y, k)$
$l(f, h)$	Luminância entre $f(x, y, k)$ e $h(x, y, k)$
$c(f, h)$	Contraste entre $f(x, y, k)$ e $h(x, y, k)$
$s(f, h)$	Medida da comparação da estrutura entre $f(x, y, k)$ e $h(x, y, k)$
$S(f, h)$	Similaridade da indexada da medida entre $f(x, y, k)$ e $h(x, y, k)$
B	Valor dinámico do <i>pixel</i>
C_1 e C_2	Constantes da SSIM
α_i	Valores da amostra α
β_i	Valores da amostra β
μ_α	Média dos valores de α
A	Número de amostras
r	Correlação entre duas medidas

CAPÍTULO 1

Introdução

O padrão ISDB-Tb (*Integrated Services Digital Broadcasting Terrestrial Built-in*), o padrão brasileiro de televisão digital e a nova tecnologia para a rede celular (3G), oferecem a TV móvel uma tecnologia que permite a transmissão de programas televisivos ou vídeo para dispositivos móveis, como telefones celulares e PDAs (*Personal Digital Assistants*). Os programas no padrão ISDB-Tb podem ser transmitidos para cada espectador em uma área, na forma de fluxo corrente (*stream*), radiodifusão ou por satélite.

Para o ISDB-Tb está prevista a recepção em vários formatos para equipamentos móveis ou fixos, com transmissão dos padrões de compressão MPEG-2 e H.264 (ALENCAR, 2007). Já para o caso do sistema UMTS (*Universal Mobile Telecommunication System*), é mostrado por Rocha *et al.* (2008) que quanto maior a taxa de *bit*, menor será a quantidade de usuários atendidos na rede.

Os vídeos usados na transmissão da TV digital possuem diferentes formatos de codificação, resoluções (espacial e temporal) e taxa de *bits*. O tipo de transmissão utilizada (a cabo, por radiodifusão, entre outras) afeta diretamente a taxa com que os *bits* podem ser enviados e consequentemente, a definição do formato mais adequado para o sinal de vídeo.

Para que os sinais de vídeo sejam utilizados adequadamente para a TV móvel são usados transcodificadores de vídeo, que são sistemas capazes de converter uma sequência de vídeo em uma outra com diferentes parâmetros, como codificação, resolução temporal e espacial e taxa de *bit*. A transcodificação também representa economia em espaço e tempo de produção, pois apenas o conteúdo da resolução máxima é armazenado.

Um sistema básico da transmissão de TV é apresentado na Figura 1.1, na qual o transcodificador pode ser implementado em três partes diferentes do sistema. A primeira parte em que ele pode ser utilizado é durante a aquisição do vídeo, depois da gravação do vídeo, gerando assim a necessidade o mesmo vídeo em diferentes resoluções, gerando uma maior espaço para armazenamento necessário para o vídeo.

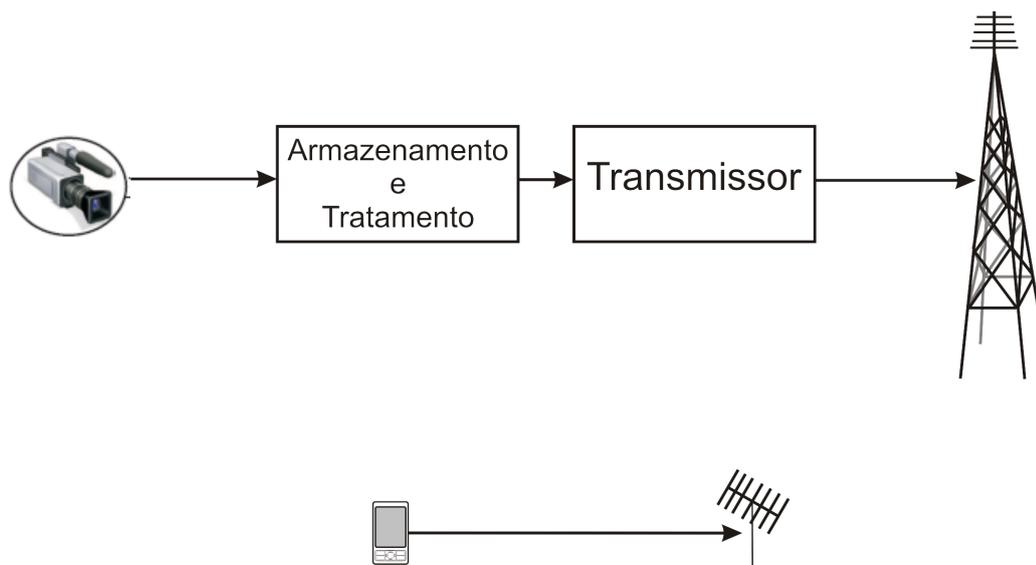


Figura 1.1 Sistema básico de TV.

A segunda parte do sistema na qual o transcodificador pode ser utilizado é em uma etapa antes da transmissão, em que se tem a vantagem de não se utilizar tanto espaço para armazenamento do vídeo.

A terceira parte do sistema em que o transcodificador pode ser utilizado no próprio receptor. Neste caso a sua utilização apresenta alguns problemas, como a necessidade de uma maior banda de transmissão necessária, já que os vídeos são maiores. Além disso, há a necessidade de se utilizar receptores com alto poder de processamento, o que em geral implica em um maior consumo de energia.

O foco deste trabalho é a segunda opção, ou seja a utilização do processo de transcodificação antes da transmissão. Um dos requisitos do transcodificador é a baixa complexidade computacional. Neste caso, para que não se produza um atraso para a recepção da TV móvel, o que é prejudicial para a prestação do serviço. Por esse motivo, são utilizadas técnicas de baixa complexidade e eficazes, para que o tempo de processamento seja o menor possível.

1.1 Objetivos

O objetivo deste trabalho é a avaliar a qualidade de vídeos transcodificados especialmente para a TV móvel e de parâmetros necessários para uma boa visualização da TV móvel. Na TV móvel, alguns parâmetros são considerados importantes para uma boa qualidade do vídeo para dispositivos móveis, como por exemplo:

- o dispositivo móvel, que nesse trabalho é avaliado o NOKIA N95, representando os celulares e o NOKIA N800, representando os PDA's;
- a taxa de *bit*, que é um limitador da transmissão por conta do limite existente.

Para a transcodificação espacial de vídeos no domínio dos *pixels* (espacial) para a TV móvel são usadas técnicas de baixa complexidade, para que o tempo de processamento seja pequeno, não atrapalhando a transmissão em tempo real. As técnicas de baixa complexidade usadas neste trabalho são os seguintes filtros passa-baixa: da média, da mediana, da moda, sigma e a média ponderada, usando três diferentes ponderações.

Cada técnica utilizada é analisada usando métricas objetivas e subjetivas com três diferentes vídeos. Os vídeos foram escolhidos por possuírem uma diversidade de conteúdo com vários graus de quantidade de movimentação, textura e cor.

1.2 Motivação

Ações efetivas para a implantação da televisão digital no Brasil começaram no final da década de 1990 com o trabalho da Comissão Brasileira de Comunicações CBC.2 da Agência Nacional de Telecomunicações (Anatel) (ALENCAR, 2005).

Em 1998, foi realizado um convênio entre a Universidade Presbiteriana Mackenzie, a Associação Brasileira de Emissoras de Rádio e Televisão (Abert) e a Sociedade de Engenheiros de Televisão (SET), com o apoio do CPqD. O objetivo era testar os sistemas de televisão digital que estavam sendo colocados em operação no mundo: o americano ASTC, o europeu DVB-T e o japonês ISDB-T. No ano de 1994, a SET e a Abert já haviam criado um grupo técnico para estudar a implantação da televisão digital no País, além de preparar empresas radiodifusoras para lidar com a nova tecnologia.

Já em 1999, a Anatel estabeleceu um termo de cooperação técnica com o CPqD e deu início ao processo de avaliação técnica e econômica para a tomada de decisão quanto ao padrão de transmissão digital a ser adotado no Brasil para o serviço de radiodifusão de sons e imagens. Assim, a Anatel autorizou as concessionárias do serviço de radiodifusão sonora de sons e imagens a executarem testes de laboratório e de campo com sistemas de transmissão digital em uso, objetivando avaliar os aspectos técnicos concernentes à melhoria da qualidade do serviço. O CPqD acompanhou os testes em campo dos três padrões existentes e prestou assessoria técnica na especificação daqueles realizados pelas concessionárias (ALENCAR, 2007).

Em 26 de novembro de 2003, foi publicado o Decreto número 4.901, instituindo oficialmente o Sistema Brasileiro de Televisão Digital (SBTV-D), composto por um Comitê de Desenvolvimento, um Comitê Consultivo (CC) e um Grupo Gestor (GG). O CPqD e a Finep foram designados como entidades de apoio técnico e administrativo, cada um com funções bem definidas, subsidiando os trabalhos do Grupo Gestor. Além disso, coube ao CPqD o papel de atuar como integrador dos projetos do sistema de TV Digital desenvolvidos pelas instituições de P&D envolvidas no projeto.

Os objetivos do projeto foram definidos, incluindo a realização de estudos elaborados pelo CPqD e por diversas instituições, contendo propostas referentes:

1. À definição do modelo de referência do Sistema Brasileiro de Televisão Digital;
2. Ao padrão de televisão digital a ser adotado no País;
3. À forma de exploração do serviço de televisão digital;
4. Ao período e modelo de transição do sistema analógico para o digital.

O decreto Presidencial número 4.901 formalizou os objetivos do SBTVD, além das composições, atribuições e competências dos comitês formados. Ficou estabelecido que o SBTVD seria financiado com recursos do Funtel. Definiram-se as seguintes premissas para a televisão digital no Brasil:

1. A razão de aspecto não deve diferir dos padrões em uso (9 x 16), para manter a compatibilidade de imagem com os outros sistemas adotados.
2. As faixas de frequências devem ser aquelas já padronizadas no País para VHF e UHF. Não é razoável propor novas faixas.
3. A programação deve ser transmitida nos sistemas SDTV e HDTV por um período de tempo determinado.
4. É importante dotar o sistema de mobilidade e para tal deve-se compensar por meio de equalização e modulação em subfaixas os efeitos de multipercurso, como a seletividade em frequência, filtragem do canal e Doppler.
5. A interatividade é requisito das geradoras, para a utilização da mídia no comércio eletrônico.
6. Os conteúdos ficam a cargo das geradoras. A interferência do governo deve ser mínima.
7. Na implementação, deve-se buscar um menor número possível de circuitos integrados, para acelerar a obtenção de fatores de escala de produção.

No projeto SBTVD foram envolvidas 105 instituições, que pesquisaram os seguintes tópicos:

1. Modulação;
2. Codificação de vídeo;
3. Camada de transporte;
4. Terminal de acesso;
5. *Middleware*;

6. Usabilidade;
7. Sincronismo de mídias;
8. Segurança da informação;
9. Canal de interatividade;
10. Transcodificação de vídeo.

O tópico de transcodificação de vídeo teve a participação do Iecom, e tratava da transcodificação de padrões de codificação, de MPEG-2 para H.264 e de MPEG-4 para H.264. A partir deste projeto surgiu a motivação para o desenvolvimento de um trabalho na área de transcodificação espacial para a TV móvel.

1.3 Contribuições

As principais contribuições apresentadas nesse trabalho são listadas a seguir:

- Comparação de aceitabilidade de vídeos visualizados em dispositivos móveis
- Qualidade da experiência do usuário em diferentes receptores móveis, como celular e PDA;
- Avaliação da qualidade dos vídeos codificados com diferentes taxas de *bits*, em dispositivos móveis;
- Avaliação dos filtros passa-baixa utilizados para a transcodificação espacial;

1.4 Estrutura do Texto

No Capítulo 2 é apresentada uma visão geral da TV móvel, apresentado os padrões existentes, com ênfase ao padrão adotado pelo Brasil. No Capítulo 3 são descritos diversos tipos de algoritmos disponíveis para a transcodificação, com ênfase na transcodificação homogênea.

No Capítulo 4 são apresentados os diversos tipos de filtros utilizados neste trabalho. No Capítulo 5 são descritos os métodos de avaliação objetivos e subjetivos considerados no trabalho. No Capítulo 6 são apresentadas as especificações do transcodificador responsável pela redução espacial dos vídeos.

A apresentação e análise dos resultados estão no Capítulo 7. As conclusões e sugestões para trabalhos futuros são apresentadas no Capítulo 8. No Apêndice A são apresentados os vários métodos de medidas subjetivas. No Apêndice B é descrito o funcionamento do codificador H.264. No Apêndice C é apresentado o questionário usado e os resultados para esse questionário. Por fim, no Apêndice D é apresentada a lista de artigos publicados pelo autor.

CAPÍTULO 2

Televisão Móvel

Televisão Móvel é a tecnologia que permite a transmissão de programas televisivos ou vídeo para dispositivos móveis, como telefones celulares e PDA. Os programas podem ser transmitidos para cada espectador em uma área na forma de *download*, por radiodifusão ou por satélite. A transmissão pode ser feita por meio terrestre, como a TV analógica ou digital é transmitida, por intermédio de satélites diretamente para os receptores móveis, ou usando a Internet e as redes celulares.

O serviço de Televisão móvel vem se desenvolvendo, com estimativa segundo a empresa norte-americana Strategy Analytics, de 51 milhões de usuários no mundo por volta de 2009, produzindo uma receita total de 6,6 bilhões de dólares (ALENCAR, 2007).

Atualmente, no telefone celular vem sendo acrescentado vários serviços, dentre eles a Televisão digital. Porém para suprir essas necessidades existem alguns itens que podem atrapalhar a continuação dessa evolução, como por exemplo: o custo elevado, a cobertura imperfeita, e a vida curta da bateria.

O uso de telefones móveis não implica que os usuários estejam em movimento, mas representa um meio de comunicação que está sempre a mão. Na televisão móvel há um compromisso entre o tamanho de tela e a portabilidade do dispositivo (CHORIANOPOULOS, 2008). As pesquisas demonstram que a maioria das pessoas interessada no serviço está disposta a aceitar um tamanho de tela menor, de forma a não comprometer a simplicidade e portabilidade do aparelho (KNOCHE; MCCARTHY, 2005), (KNOCHE; MCCARTHY, 2004). Entretanto, muitas pessoas optariam por um dispositivo separado que tivesse uma tela maior deixando o aparelho celular dedicado a inter-comunicação. Em ambos os casos, a qualidade do vídeo é um fator importante para o sucesso da TV móvel.

As maiores diferenças entre a transmissão para televisores tradicionais e para receptores móveis são: os receptores móveis possuem uma largura de faixa limitada, o que implica em maiores exigências de compressão e severas limitações com relação ao tamanho do vídeo, os recursos de processamento e a memória disponível. A produção do vídeo que atenda a essas exigências depende da sua qualidade. Esta qualidade, por sua vez, é afetada pelas distorções

inseridas pelo sistema de comunicações (canal, display, aquisição). Por essa razão, a fim de assegurar uma qualidade de serviço aceitável, testes subjetivos da avaliação de qualidade durante o desenvolvimento do produto são necessários (CHORIANOPOULOS, 2008).

2.1 Transmissão da Televisão Móvel

Existem três abordagens para a transmissão do conteúdo para receptores móveis, o modo por radiodifusão, o modo *multicast* e o modo *unicast*. No modo por radiodifusão o conteúdo é disponibilizado para um número ilimitado de usuários, ideal para a transmissão de canais de TV com demandas universais. O modo *unicast*, por outro lado, é projetado para fornecer ao usuário vídeos selecionados ou outros tipos de serviços de áudio/vídeo. A conexão virtual é diferente para cada usuário, com o usuário selecionando o conteúdo a ser transmitido, como também os outros serviços de interatividade. O modo *unicast* tem limitações no número de usuários que podem usar os recursos. Isto funciona bem para redes 3G, uma vez que a infra-estrutura já está disponível. Por exemplo, o fluxo de vídeo de um programa esportivo pode ser selecionado por centenas de milhares de usuários, resultando no esgotamento dos recursos para transmissão de tais serviços. Ao mesmo tempo, o grau de utilização de serviços específicos que podem ser fornecidos é muito elevado. Na Figura 2.1 são apresentadas os modos de transmissão por radiodifusão e *unicast*.

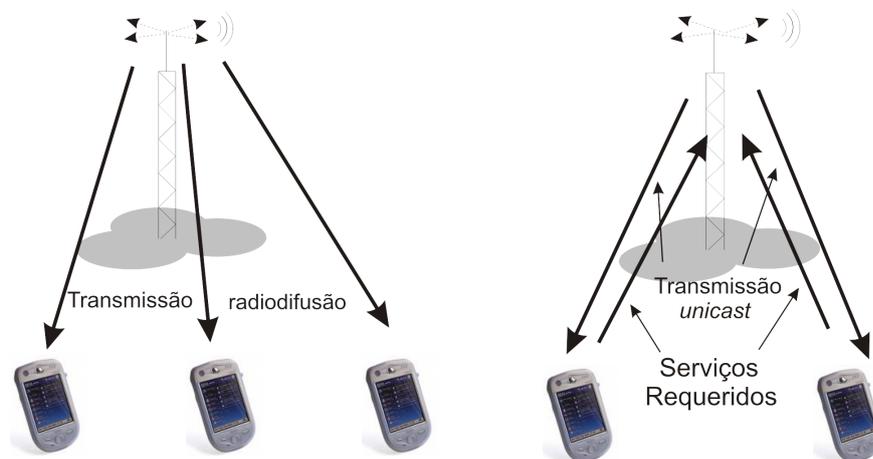


Figura 2.1 Transmissão para TV móvel por radiodifusão e *unicast*.

No *multicast* o mesmo vídeo é transmitido a partir de uma única fonte para grupos selecionados de usuários (grupo *multicast*). Um grupo *multicast* poderia ter usuários com interesses similares, tais como esportes ou notícias. Exemplos de tecnologias de *multicast* são MBMS e o WiMAX móvel, que pode ser operado em *unicast* também.

As transmissões *multicast* e por radiodifusão são muito parecidas, e suas diferenças são:

1. O modo *multicast* envolve a transmissão, desde o início até todos os dispositivos de um grupo. Esses dispositivos podem residir em diferentes áreas da célula ou serem móveis. A transmissão *multicast* não é feita a todos os destinatários de uma área.
2. A modalidade por radiodifusão transmite dados para todos os usuários de uma área.

2.1.1 Televisão Móvel em Redes Celulares

Os receptores para TV móvel são aparelhos que estão ligados a redes 2G e 3G. Os aparelhos telefônicos podem ser utilizados para comunicações de voz e dados, além de servir como receptores para TV Móvel.

As operadoras querem difundir tais redes, por reconhecerem que existe uma capacidade para o canal de retorno na comunicação no celular, o que pode ajudar a proporcionar aplicações interativas. O WiMAX, WiBro, e as redes LANs sem fio são opções para canais de interatividade.

As operadoras móveis oferecem *download* de vídeo e áudio desde o início da tecnologia 2.5G. Os vídeos disponíveis para *download* são geralmente de curta duração (poucos segundos), devido à baixa taxa de dados da rede.

Como as redes estão mudando para o 3G, as taxas de dados aumentaram e tem protocolos definidos para transmissão de áudio e vídeo. Por esta razão, aumentou a oferta de vídeos pelos canais de 3G, com taxa de *bit* de 128 *kbit/s* ou mais, para o enlace de subida, enquanto para o enlace de descida é usada 12,2 *kbit/s* (ROCHA, 2008). Com a codificação MPEG-4 ou H.264 é possível transmitir vídeos com estas taxas.

O sistema UMTS que apresenta características de padrões 3G exigidas pela ITU e pelo IMT-2000, divide as altas taxas de transmissão em três categorias (HOLMA; TOSKALA, 2004; ROCHA, 2008):

- 2 *Mbit/s* em cenários em que o equipamento do usuário se movimenta pouco, ou ambientes internos;
- 384 *kbit/s* em cenários onde os usuários são pedestres ou ambientes urbanos;
- 144 *kbit/s* em ambientes com muito movimento do usuário;
- Taxas de dados variáveis em sistemas com cobertura de grande área geográfica (satélites);

2.1.2 Televisão Móvel em Redes de Televisão Digital

O conceito de TV móvel usando redes terrestres é algo semelhante ao dos receptores de rádio FM embutidos nos aparelhos celulares. O rádio funciona a partir da recepção dos canais FM e não utiliza da capacidade das redes 2G ou 3G. Os aparelhos têm um sintonizador embutido e um demodulador para os sinais FM. Mesmo se não houver cobertura das redes de 2G ou 3G

o rádio FM continuará a funcionar. A TV móvel utilizando tecnologias de transmissão terrestre segue o mesmo conceito da televisão analógica e utiliza o espectro de VHF ou UHF.

2.1.3 Parâmetros de Avaliação dos Padrões

Pode-se citar alguns parâmetros importantes na avaliação das tecnologias, já que existem vários sistemas de transmissão, como *unicast*, *multicast* e radiodifusão:

- robustez da transmissão e da qualidade do serviço em ambientes internos e ao ar livre;
- características de economia de energia;
- troca de canal no tempo;
- aparelhos necessários para apoiar o serviço;
- utilização eficiente do espectro;
- custos de exploração de serviços;
- características, tais como a qualidade, custos e recepção. Essas características, dependem das redes subjacentes (por exemplo, 3G);
- as necessidades do usuário, tais como a disponibilidade no país, a capacidade de *roaming*, tipos de aparelhos telefônicos, e serviços disponíveis.

2.2 Padrões de Televisão Móvel

No momento, as operadoras de telefonia móvel estão fornecendo serviços de vídeo usando suas redes UMTS, ou redes de terceira-geração. Os serviços de televisão ao vivo em um dispositivo portátil estão atualmente disponíveis em DMB (*Digital Multimedia Broadcast*), ISDB-T, MediaFLO, DVB-H (GROUP, 2005), (KUMAR, 2007) e DVB-SH (BROADCASTING, 2007).

O DVB-H é uma tecnologia desenvolvida como extensão do DVB-T, com alguns recursos adicionais que a torna apropriada para a transmissão para dispositivos móveis. Os dispositivos móveis, como celulares e PDA, são operados com bateria, com necessidade de conservar energia, com pequenas antenas de baixo ganho, e têm a necessidade de manter a complexidade baixa no receptor.

A tecnologia do DVB-H vem sendo vista como muito promissora, pois pode efetivamente utilizar a infra-estrutura do DVB-T e o espectro disponível, fornecendo uma transmissão de Televisão de qualidade ao número de usuários da região de cobertura (FARIA *et al.*, 2006).

O DMB fornece serviço de televisão móvel usando o padrão Eureka-147 DAB (*Digital Audio Broadcasting*) com correção adicional de erro. A nova norma para o DMB foi formalizada pelo ETSI (*European Telecommunications Standards Institute*) como ETSI TS 102428. O DMB tem a opção de transmissão por satélite (DMB-S) ou uma opção de transmissão terrestre (T-DMB), utilizada na Coreia e no Japão. A opção T-DMB usa a rede terrestre na banda III ou na banda L, enquanto que o S-DMB usa a rede via satélite na banda L (BROOKS, 2005).

O ISDB, desenvolvido pelo Japão como seu padrão de televisão digital terrestre, tem a sua banda dividida em 14 segmentos, cada segmento com uma largura de banda de 433 kHz, sendo que um dos segmentos é utilizado como intervalo de guarda. Os treze segmentos restantes são utilizados para a transmissão de TV para dispositivos móveis, por isso o padrão de TV Móvel é chamado de *One-Seg* (TAKADA; SAITO, 2006).

MediaFLO é um sistema proprietário desenvolvido pela Qualcomm para transmissão de serviços para receptores portáteis usando OFDM. A Qualcomm oferece esse tipo de serviço na faixa de frequência de 700 MHz, nos Estados Unidos, uma vez que detém uma licença nesta parte do espectro (KUMAR, 2007). Nos Estados Unidos, as operadoras AT&T e Verizon oferecem conteúdos via MediaFlo para seus clientes, em sete modelos de aparelhos (ADNEWS, 2008).

O DVB-H ou ISDB-T baseiam-se na transmissão terrestre aproveitando as redes já existentes. Outros optaram por usar redes terrestres inteiramente novas ou via satélite para a Televisão Móvel. A Televisão Móvel que utiliza transmissão terrestre é uma classe importante, devido à elevada potência que os transmissores terrestres podem proporcionar atingindo os receptores móveis, mesmo com pequenas antenas internas e longas distâncias da antena. Na Tabela 2.1 está resumido a transmissão terrestre de Televisão móvel baseada nas tecnologias de transmissão por radiodifusão.

Tabela 2.1 Comparação das tecnologias de Televisão Móvel baseadas em transmissão por radiodifusão

Características	DVB-H	T-DMB	ISDB-T
Formatos de Vídeo e Áudio	vídeo MPEG-4 ou WM9 áudio AAC ou WM	vídeo MPEG-4 áudio BSAC	vídeo MPEG-4 áudio AAC
Fluxo de Transporte	IP sobre MPEG-2 TS	MPEG-2 TS	MPEG-2 TS
Modulação	QPSK ou 16 QAM com COFDM	DQPSK com FDM	DQPSK ou QPSK ou 16-QAM ou 64-QAM com BST-OFDM
Largura de Banda RF	5-8 MHz	1.54 MHz (Coreia)	433 kHz (Japão)
Economia de Energia	Divisão de tempo	Redução da banda	Redução da banda

Este trabalho apresenta o estudo do ISDB-T, e especificamente o *One-Seg*, que é o padrão para Televisão móvel adotado pelo Brasil.

2.3 One-Seg

O ISDB foi inicialmente implementado no Japão em 2000 sob a forma de ISDB-S (Satélite). O sistema ISDB-T utiliza a transmissão com a modulação OFDM de banda segmentada (BST-OFDM), que permite serviços de HDTV e serviço de recepção móvel utilizando a transmissão por radiodifusão (PAIVA *et al.*, 2008).

O DTTB (*Digital Terrestrial Broadcasting*) começou no Japão em Dezembro 2003 e, desde então, foi progressivamente substituindo as transmissões analógicas (NTSC). A transmissão consiste de canais de 6 MHz. A maioria das transmissões do DTTB é feita agora em HDTV.

A Televisão Móvel com ISDB-T começou no Japão em 2006. Os parâmetros de vídeo e áudio codificados para o ISDB-T são:

- Vídeo – codificado utilizando H.264 *base line profile* L1.2 com 15 quadros/s e resolução QVGA (320 x 240);
- Áudio – MPEG-2 AAC com amostragem de 24,48 k amostras/s.

O serviço de recepção móvel utiliza parâmetros de modulação robusta, e tem banda muito limitada, em comparação com o serviço HDTV. A taxa de transmissão total varia muito, dependendo dos parâmetros da transmissão, sendo 320 *kbit/s* um valor considerado adequado por (KUMAR, 2007) para a recepção móvel. Na Figura 2.2 é apresentado uma ilustração da divisão de bandas no ISDB-T.

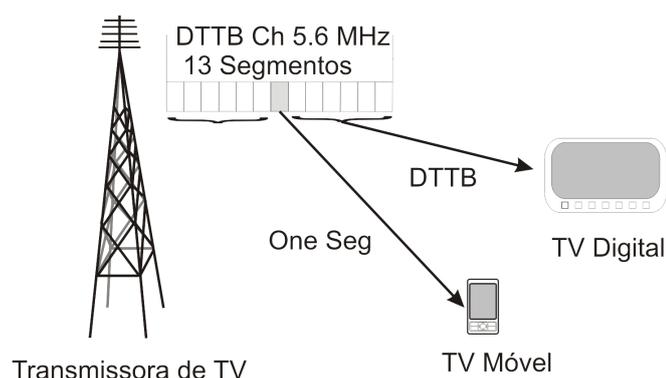


Figura 2.2 Serviço ISDB-T.

A televisão digital no Brasil teve sua primeira transmissão oficial em 2 de dezembro de 2007, às 21h20, na cidade de São Paulo. A solenidade reuniu mais de 2000 pessoas e contou com a presença do Presidente da República Luiz Inácio Lula da Silva e de grandes empresários do setor.

2.3.1 Parâmetros de Transmissão

Em (MATSUMURA *et al.*, 2005) foram apresentados os resultados de testes para a transmissão. Na Tabela 2.2 são apresentados os componentes e a taxa distribuição dos serviços utilizados nos testes. A taxa total de transmissão derivada da modulação dos parâmetros foi 312 *kbits/s*. A taxa real média total foi de aproximadamente 300 *kbits/s* considerando flutuações nos fluxos de vídeo e áudio. A taxa de PSI/SI (Serviço de informações de sincronização do relógio, multiplexador de parâmetros, nome do programa, etc), foi decidida a partir da composição do serviço. Em geral, a solução de compromisso entre qualidade do vídeo/áudio e a velocidade de transmissão exigida para transmissão dos dados determina a taxa de distribuição do vídeo, áudio e dados.

Tabela 2.2 Parâmetros de modulação: QPSK, intervalo de guarda de 1/8, mode3, corretor de erro convolucional de 1/2.

Elemento	Parâmetros de codificação e transmissão	taxa média (<i>kbits/s</i>)
Vídeo	MPEG-4, 240x136, 15 quadros/s	176
Audio	MPEG-2 AAC 24 kHz estéreo	52
PSI/SI	PCR, NIT,SDT, PMT, L-EIT	29,25
Dados de radiodifusão	MPEG-2 DSM-CC	40
Total		297,25 (312)

2.4 Televisão Móvel no Brasil

O sistema brasileiro de televisão digital, que recebeu, em um primeiro momento, como nomenclatura oficial a sigla SBTVD (Sistema Brasileiro de Televisão Digital), teve seu nome alterado posteriormente para ISDTV (*International System for Digital Television*) e finalmente, nos dias de hoje, tem como sua denominação oficial o termo ISDB-Tb (*Integrated Services Digital Broadcasting Terrestrial Built-in*).

Esse sistema foi desenvolvido para prover transmissão confiável de vídeo, áudio e dados de alta qualidade não só para receptores fixos, mas também para receptores móveis. O sistema ISDB-Tb tem a capacidade de oferecer flexibilidade, rápida capacidade de expansão e interoperabilidade às emissoras de conteúdo digital. O ISDB-Tb herdou praticamente todas as características do padrão japonês ISDB-T no que diz respeito as camadas de transmissão, codificação de canal e codificação de fonte. Uma das principais mudanças foi a adoção do padrão de decodificação de vídeo digital ITU-T H.264 (OLIVEIRA, 2008).

Esta seção discute o padrão de Televisão Móvel que será utilizado no Brasil. Destaca-se o receptor, a codificação de vídeo e áudio e o mercado atual. São vistos na codificação de vídeo os parâmetros e as resoluções necessárias para a transmissão da TV Móvel.

2.4.1 Receptor de Televisão Móvel

A classificação *One-Seg* é destinada aos receptores do tipo portátil, também conhecidos por *handheld*, especialmente recomendados para telas de exibição de dimensões reduzidas, normalmente até sete polegadas. Dentre os produtos classificados como *One-Seg*, estão os receptores integrados com telefone celular, PDA e televisores portáteis, que são energizados por uma bateria interna e, portanto, sem necessariamente demandar uma fonte externa de energia, bem como aqueles destinados a veículos automóveis. Esse tipo de receptor é capaz de receber e decodificar apenas sinais de televisão digital terrestre transportados na camada "A" do fluxo de transporte, e, conseqüentemente, apenas sinais de perfil básico, destinados aos dispositivos portáteis de recepção (ABNT, 2007a).

A configuração básica do receptor está na Figura 2.3 e é composta pelas seguintes unidades:

- a) antena de recepção terrestre;
- b) IRD (*Integrated Receiver Decoder*);
- c) cabo de conexão entre a antena e o receptor.

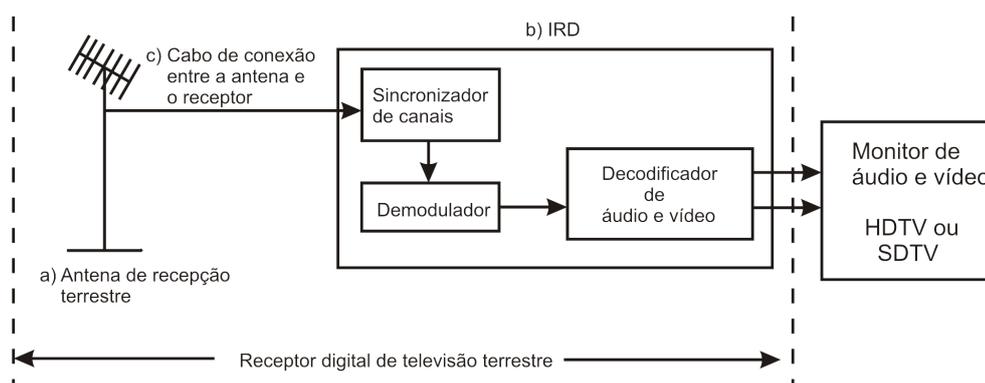


Figura 2.3 Configuração básica do receptor.

É recomendado que esse tipo de receptor execute buscas automáticas de canais periodicamente em ciclos de tempo que podem ser definidos pelo fabricante do produto.

Para a interatividade o receptor móvel deve ter o *middleware* embarcado, e deve obrigatoriamente contemplar as especificações do Gingal-NCL *stand alone* com máquina de execução LUA. A ponte com uma máquina Java é opcional (ABNT, 2007b).

2.4.2 Codificação de Vídeo

A ferramenta de compressão de vídeo do sistema de televisão digital terrestre brasileiro é o H.264 e deve estar de acordo com a ITU-T (*International Telecommunication Union*) Recommendation H.264 (ABNT, 2007b). Essa é umas das principais diferenças para o padrão japonês, já que o esse utiliza o codificador MPEG-2. A principal vantagem do H.264 para o MPEG-2 é a maior compressão dos vídeos com uma perda de qualidade menor.

Na codificação do vídeo existem vários perfis, que estabelecem o nível de complexidade. Os perfis definem um subconjunto da sintaxe do fluxo de *bits* que é adotado, visando, em geral, determinadas classes de aplicações. Além dos perfis existem os níveis, que estabelecem para cada perfil os valores permitidos de alguns parâmetros de codificação como, por exemplo, taxas de *bits* e resolução.

Os níveis acrescentam restrições aos perfis, para garantir a interoperabilidade entre os diferentes decodificadores, especificando o limite de desempenho requerido para realizar a decodificação.

A norma estabelece a utilização das restrições para o One-seg dadas na Tabela 2.3. O perfil recomendado é o *baseline*. Dentre uma das suas características tem-se que o vídeo codificado só deve conter *slices* do tipo I ou P. Como pode ser visto na Tabela 2.3, o nível pode utilizar até o limite de 1.3, dependendo da aplicação. Na Tabela 2.4 são apresentadas as limitações para cada nível.

Tabela 2.3 Restrições aos parâmetros de codificação.

Taxa de <i>bits</i>	64 <i>kbit/s</i> até a máxima taxa de <i>bits</i> permitida pelo perfil@nível especificado na ITU-T Recommendation H.264:2005
Resolução de imagem	SQVGA (160 x 120 ou 160 x 90), QVGA (320 x 240 ou 320 x 180) e CIF (352 x 288)
Taxas de quadros	5, 10, 12, 15, 24, 30 quadros/s
Razão de aspecto do <i>display</i>	4:3 ou 16:9
Nível	Até o limite de 1.3, dependendo da aplicação
Outras restrições	FMO (<i>Flexible Macroblock Ordering</i>), ASO (<i>Arbitrary Slice Ordering</i>) e RS (<i>Redundant Slices</i>) não podem ser utilizados

2.4.3 Resoluções de Luminância

São permitidas as resoluções até o limite imposto pelo nível 1.3, que restringe o número máximo de macroblocos para cada quadro em 396. Esse número corresponde tipicamente à resolução máxima de 352 x 288 (CIF) para a luminância. Sendo possível o uso dos seguintes formatos:

Tabela 2.4 Limitações impostas por cada nível.

Número do nível	Taxa máxima de processamento de macroblocos MaxMBPS MB/s	Tamanho máximo do quadro MaxFS MBe	Tamanho máximo do buffer de decodificação MaxDPB(1024 bytes para 4:2:0)	Taxa máxima de vídeo MaxBR (1000 <i>bits/s</i> , 1200 <i>bits/s</i> , <i>cpbBrVclFactor bits/s</i> , ou <i>cpbBrNalFactor bits/s</i>)	Tamanho máximo CPB (1000 <i>bits</i> , 1200 <i>bits</i> , <i>cpbBrVclFactor bits/s</i> , ou <i>cpbBrNalFactor bits/s</i>)
1	1485	99	148,5	64	175
1b	1485	99	148,5	128	350
1.1	3000	396	337,5	192	500
1.2	6000	396	891,0	384	1000
1.3	11800	396	891,0	768	2000

- SQVGA, com resolução 160×120 e razão de aspecto de 4:3;
- SQVGA, com resolução 160×90 e razão de aspecto de 16:9;
- QVGA, com resolução 320×240 e razão de aspecto de 4:3;
- QVGA, com resolução 320×180 e razão de aspecto de 16:9;
- CIF, com resolução 352×288 e razão de aspecto de 4:3.

Se o número de amostras de luminância para a largura e a altura não for múltiplo de 16, amostras extras devem ser acrescentadas à imagem, de forma a atingir esse múltiplo. Na largura, recomenda-se que as amostras extras sejam adicionadas no lado direito da imagem e, para a altura, recomenda-se que as amostras extras sejam adicionadas na parte inferior da imagem.

2.4.4 Mercado Atual Brasileiro

Com o início da TV digital aberta no dia 2 de dezembro 2008 na capital de São Paulo foi aberto um importante meio para as transmissões da TV Móvel. Porém esse tipo de serviço tem alguns problemas, dentre eles figura a disputa entre as empresas de radiodifusão e as de telecomunicações por esse mercado. Há também a questão da capacidade dos próprios aparelhos para a captação dos sinais de Televisão Móvel.

Em algumas partes do Brasil já esta sendo usada a Televisão Móvel via redes celulares. Esse serviço pode ser contratado por duas horas ou 24 horas, com custos respectivos de R\$ 5,90 reais e R\$ 9,90 reais. Para acessá-lo, é necessário ter um aparelho compatível com a rede EDGE (MOREIRA., 2007).

A primeira novela brasileira feita exclusivamente para celulares foi Parafina, exibida em fevereiro de 2007, apenas para clientes da Oi. Foram 24 capítulos de uma história sobre surfistas filmada no Rio de Janeiro, distribuída três vezes por semana em *downloads* pagos via

WAP. A recepção do público foi tão boa que a operadora lançou em 2007 um canal de televisão via celular o Oi Televisão Móvel, primeiro do gênero no mercado brasileiro (ABERT, 2009).

O primeiro aparelho compatível com o sistema brasileiro de Televisão digital foi lançado pela Samsung no dia quatro de março de 2008. Além de receber o sinal de Televisão aberta digital, o aparelho funciona em redes de terceira geração (3G) e está pronto para fazer chamadas de vídeo. A duração da bateria é de até 3,5 horas, no modo recepção de Televisão. O modelo V820L (Figura 2.4) tem *display* OLED colorido, câmera com resolução de 2 *Mpixels*, MP3 *player*, memória expansível via cartão microSD e conexão *Bluetooth* e chega ao mercado com preço sugerido de 1.599 reais a 1.699 reais (SAMSUNG, 2008).



Figura 2.4 Aparelho V820L, para transmissão de TV Móvel (SAMSUNG, 2008).

2.5 Aceitabilidade e Avaliação da Qualidade Visual

O serviço de TV móvel prevê um usuário com diferentes expectativas, que é uma grande mudança em relação à Televisão tradicional. Embora o serviço de Televisão móvel tenha a vantagem do conteúdo estar em todo lugar, têm a desvantagem da tela com dimensão pequena, da largura de banda da transmissão baixa, altas taxas de erro, e pequena cobertura. O êxito do serviço de Televisão móvel depende da sua capacidade de satisfazer as expectativas dos usuários e proporcionar uma boa experiência (AHMAD; IBRAHIM, 2008).

Os resultados mostrados por Knoche *et al.* (2008) indicam que o objeto da aceitabilidade a Televisão móvel depende do tipo de conteúdo, bem como da resolução em que o conteúdo será exibido. A escolha de conteúdos de qualidade é um fator chave na determinação do êxito do serviço de televisão móvel.

Exemplos de conteúdos que têm a preferência dos usuários são notícias, filmes, comédia, esportes e viagens. A questão sobre a duração do conteúdo ainda não está bem definida devido ao problema da bateria e da aceitabilidade do usuário em assistir em uma tela pequena. Estudos efetuados por Knoche *et al.* (2005) sugerem que o tempo de uso seja muito curto, geralmente dentro de 2 a 5 minutos.

A integração da qualidade do áudio e vídeo depende do conteúdo, pois em cenas com pouca movimentação (por exemplo, jornais), a qualidade do áudio é mais importante do que a qualidade de vídeo. Assim, para vídeos com muito movimento, um controle da qualidade mais rigoroso é necessário para que os usuários tenham uma boa aceitação (HANDS, 2004).

CAPÍTULO 3

Sistema de Transcodificação

Na literatura, há vários trabalhos sobre os tipos de transcodificação e as formas de se realizar cada uma delas, como mostrado por Ahmad *et al.* (2005), Xin *et al.* (2005) e Vetro *et al.* (2003).

Para contextualizar os objetivos deste trabalho, nesta subseção são apresentadas algumas contribuições de trabalhos anteriores.

Alguns trabalhos apresentam formas de transcodificação da redução espacial com o uso de vetores de movimento, como mostrado por Xin *et al.* (2005), Vetro *et al.* (2003), Xu *et al.* (2004). Entretanto, essa técnica não foi utilizada neste trabalho devido ao elevado nível de complexidade. Um outro método que foi descartado pelo elevado nível de complexidade foi a transcodificação da redução espacial usando a DCT, como mostrado por Lee *et al.* (2004) e Yin *et al.* (2000).

As referências (XIN *et al.*, 2002) e (QUEIROZ, 2005) mostram um outro método que pode ser realizado para a transcodificação espacial, como ilustrado na Figura 3.1. Como esse trabalho trata de vídeos no domínio dos *pixels* a Figura 3.1 pode ser simplificada, pelos blocos que se encontram dentro área pontilhada da figura.

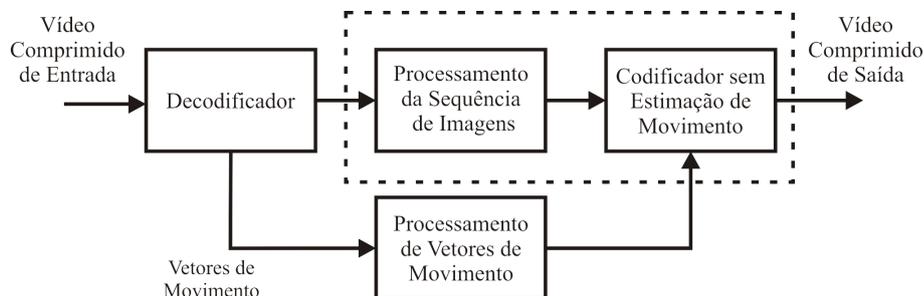


Figura 3.1 Modelo de um transcodificador de vídeo.

A referência (XIN *et al.*, 2002) apresenta detalhadamente o bloco processamento das sequências de imagens da Figura 3.1, como apresentado na Figura 3.2. Esse processo se mostra eficaz principalmente por ter um tempo de execução baixo, se comparado como os outros

métodos da transcodificação da resolução espacial de vídeos. O processamento das sequências de imagens é realizado por um interpolador fracional na qual L e S significam os fatores de aumento e diminuição espacial, com taxa L/S.

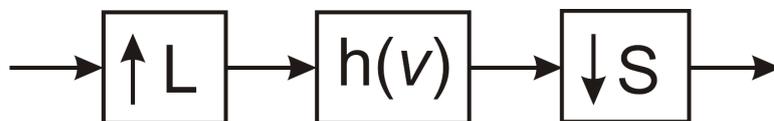


Figura 3.2 Interpolador fracional com taxa de L/S.

O bloco $h(v)$ é um filtro passa-baixa, do tipo da média, da mediana, da moda e sigma. Esses filtros são amplamente encontrados na literatura para filtragens de imagens e apresentam baixa complexidade (ARAÚJO, 1987), (PITAS; VENETSANOPOULOS, 1986), (BOVIK *et al.*, 1983) e (GONÇALVES; GARCIA, 1996).

Em (XIN *et al.*, 2002) é feita a transcodificação espacial de HDTV para SDTV usando vetores de movimento. Para a TV móvel é interessante que a transcodificação seja feita de SDTV (720×576) para CIF (352×288), porém é mostrado em (REGIS *et al.*, 2007) que para fazer esta transcodificação usa-se o fator $L = 1$ e $S = 2$ que é o mesmo fator de redução do CIF para QCIF (176×144). Sabendo que os fatores L e S são os mesmos para ambas as direções (vertical e horizontal), foi escolhida para esse trabalho a redução de CIF para QCIF por existirem vídeos padrões em ambos os formatos, o que é necessário para fazer os testes objetivos.

3.1 Tipos de Transcodificação

O transcodificador converte uma sequência de vídeo em outra sequência com diferentes parâmetros. Sintaxe, resolução espacial, resolução temporal, taxa de *bits* ou até mesmo a adição de funcionalidades, são exemplos de parâmetros usados.

A transcodificação pode ser classificada como homogênea, heterogênea ou com funções adicionais. A transcodificação homogênea modifica um conjunto de parâmetros do vídeo sem modificar o seu padrão de compressão, ou seja, realiza conversão da taxa de *bits*, da resolução espacial, da resolução temporal e mudanças da codificação VBR (*Variable Bit-Rate*) para CBR (*Constant Bit-Rate*). A transcodificação heterogênea realiza a conversão de padrões de compressão ou a conversão entre os formatos entrelaçado e progressivo. Além desses dois tipos, um outro tipo de transcodificação consiste em incluir funções adicionais à sequência de vídeo codificada, tais como proporcionar resistência a erros (*error resilience*) e adicionar logomarcas ou marcas d'água invisíveis (AHMAD *et al.*, 2005). Um diagrama com os vários tipos de transcodificação é apresentado na Figura 3.3 (AHMAD *et al.*, 2005; XIN *et al.*, 2005, 2002).

Nesta seção, descreve-se os tipos de transcodificação que fazem parte do esquema proposto neste trabalho: a transcodificação homogênea (resolução espacial, resolução temporal e taxa de *bits*) e a heterogênea (mudança de formato e de compressão).

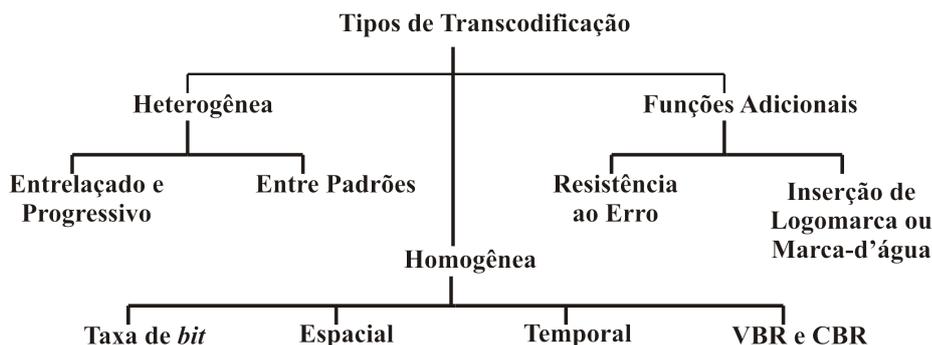


Figura 3.3 Diagrama dos possíveis tipos de transcodificação.

3.1.1 Transcodificação Homogênea

Resolução Espacial

O objetivo na transcodificação da resolução espacial é reduzir ou aumentar a resolução espacial do vídeo, utilizando técnicas como subamostragem/interpolação no domínio dos *pixels*.

A interpolação, subamostragem e filtragem são técnicas para reduzir ou aumentar a definição espacial. O interpolador é empregado para aumentar a taxa de amostragem de uma sequência por um fator inteiro $L > 1$, enquanto o subamostrador é empregado para diminuir a taxa de amostragem por um fator inteiro $S > 1$ (VAIDYANATHAN, 1993), (MITRA, 2006).

Usa-se o interpolador e o subamostrador juntos quando é necessário a redução ou aumento por um fator não inteiro, como mostrado na Figura 3.2. Na próxima seção são mostrados os tipos de filtros possíveis.

Resolução Temporal

O objetivo principal da redução da resolução temporal, que consiste na redução do número de quadros por segundo. Essa redução é importante porque uma maior taxa de quadros requer uma maior largura de banda necessária para transmissão. Além disso, tal diminuição provoca uma economia de bateria nos dispositivos móveis, por diminuir a carga computacional imposta ao aparelho para reprodução dos vídeos, como mostrado por Vetro *et al.* (2003).

A técnica também pode ser usada para redução de taxa de *bits* imposta pelas redes de comunicações e para manter a qualidade dos quadros em situações em que há limitação de processamento imposta pelo equipamento do consumidor.

Como descrito por (AHMAD *et al.*, 2005), a maioria das técnicas de redução temporal utiliza vetores de movimento. Tal técnica não é descrita nesse trabalho, uma vez que o sistema em questão realiza a transcodificação de vídeo no domínio dos *pixels*.

Ajuste da Taxa de *Bit* de Vídeo

A transcodificação de taxa de *bits* tem como objetivo reduzir a taxa de *bits* do vídeo codificado, mantendo uma baixa complexidade, por se ter menos dados a processar. Requantização e transmissão seletiva são exemplos de técnicas utilizadas para alcançar tal objetivo.

Na requantização, existe uma boa relação entre complexidade e qualidade da imagem, tornando possível o controle da redução da taxa de *bits*. Essa técnica consiste em aumentar o grau de quantização no codificador do transcodificador, diminuindo o número de coeficientes não nulos quantizados e conseqüentemente, a dimensão do vídeo codificado.

Na transmissão seletiva eliminam-se alguns coeficientes de altas frequências, preservando, ainda assim, a qualidade da imagem, uma vez que a maior parte da energia do sinal de vídeo é concentrada na banda de baixas frequências. Essa técnica pode introduzir efeito de blocagem no vídeo resultante (SHANABLEH; GHANBARI, 2000).

3.1.2 Transcodificação Heterogênea

Existem dois tipos de transcodificação heterogênea: mudança de padrões de compressão e mudança do formato entrelaçado para progressivo (AHMAD *et al.*, 2005), (POYNTON, 2003). Na transcodificação heterogênea pode-se adicionar funcionalidades da transcodificação homogênea, assim como suas técnicas.

Como os sinais de vídeo apresentam diferentes formatos de codificação, a mudança entre padrões torna-se importante e muito utilizada, uma vez que é necessária em alguns receptores. Dentre esses padrões podem ser citados o H.263, MPEG-2, MPEG-4 e H.264.

Os quadros dos vídeos entrelaçados são formados por dois campos distintos, chamados de campos de linhas pares e campos de linhas ímpares. Esse tipo de vídeo possui uma taxa de quadros entre 50 e 60 quadros por segundo e resolução espacial (número de linhas) correspondente à metade da resolução espacial do quadro. Nos vídeos progressivos os quadros possuem um único campo, com taxa de quadros entre 25 e 30 quadros por segundo (POYNTON, 2003).

Em (XIN *et al.*, 2005) há um exemplo de transcodificação de padrões, de MPEG-2 para MPEG-4 Simple Profile (SP). O MPEG-4 (SP) é utilizado em aplicações de vídeo de baixa complexidade e de baixa taxa de *bits*, comparado com o MPEG-2. O processo de conversão utiliza duas etapas, a primeira etapa é responsável pela redução da taxa de quadros e do tipo de quadro da conversão. A segunda etapa é responsável pela redução da resolução espacial e da mudança do formato entrelaçado para o progressivo.

3.2 Filtros no Domínio Espacial

A filtragem aplicada a um vídeo digital é uma operação local que modifica os valores dos níveis de cada *pixel* do quadro, considerando o seu contexto atual. Essa técnica é uma transformação do quadro *pixel a pixel*, que depende do valor de um determinado *pixel* e do

valor dos seus vizinhos, no quadro original, ou seja, o *pixel* filtrado tem um valor dependente do contexto em que ele se encontra no quadro original.

3.2.1 Princípio de Funcionamento

O princípio de funcionamento dos filtros que operam no domínio espacial baseia-se em relações de vizinhança entre os elementos de uma região de tamanho e formato predeterminado. Domínio espacial refere-se ao plano do próprio quadro, sendo que nesta categoria trabalha-se diretamente com o valor dos *pixels* de um quadro. No processo de filtragem são atribuídos valores aos elementos do quadro destino em função dos elementos presentes no quadro original. Numa classificação mais detalhada, as técnicas encontradas na literatura podem ser divididas em quatro grupos (ARAÚJO, 1987):

- Suavização espacial de regiões, que emprega filtros passa-baixas lineares e não lineares no domínio espacial, tem como finalidade a remoção de ruído e uniformização dos níveis de cinza dos *pixels*;
- Reescalamento dos níveis de cinza, que manipula ou requantiza os níveis de cinza para aumentar o contraste;
- Realce de bordas, que envolve filtros passa-altas lineares e não lineares no domínio espacial;

Na filtragem é possível utilizar formatos diferentes e valores dependentes da posição no quadro. Esses formatos são conhecidos como máscara de deslocamento ou janela móvel. Na Figura 3.4, é mostrada uma máscara com dimensões horizontais e verticais idênticas e iguais a três.

O elemento central da máscara, definido pelas coordenadas (x,y) do quadro recebe um valor calculado em função de todos os elementos do quadro pertencentes à máscara utilizada. No caso mostrado, esses elementos variam na horizontal de $(x + 1)$ a $(x - 1)$ e na vertical de $(y + 1)$ a $(y - 1)$, resultando num total de nove elementos por máscara.

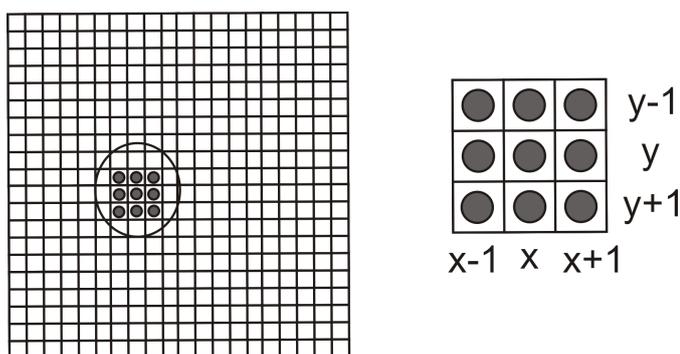


Figura 3.4 Máscara de deslocamento.

O formato da máscara de deslocamento mais comumente utilizado é o quadrado. Dependendo do algoritmo, pode-se ter submáscaras com outros formatos, como triangulares, pentagonais ou formatos irregulares. Até mesmo a máscara principal, que na maioria dos casos é a única, pode ter formatos diferentes, como a cruz ou retângulo. Alguns algoritmos admitem a escolha do formato da máscara de deslocamento e outros são projetados para um tipo específico de máscara, existindo também algoritmos em que o tamanho da máscara é predeterminado.

Geralmente, o deslocamento da máscara é realizado da esquerda para a direita sobre cada linha do quadro. Finalizando a linha, o centro da máscara é posicionado no início da linha seguinte, e o processo é repetido até que a máscara alcance o canto inferior direito do quadro.

Como pode ser visto, a utilização desse procedimento causa um problema nas bordas do quadro, pois a máscara deve estar com o seu centro sobre a posição (x, y) que se deseja calcular. Com a máscara centralizada em um ponto da borda, alguns elementos da máscara não terão valores definidos por estarem fora dos limites do quadro. Para contornar este problema, pode-se restringir o posicionamento da máscara para que ela não seja sobreposta a pontos não pertencentes ao quadro ou atribui-se aos pontos fora do quadro um valor predeterminado. No primeiro caso, após a execução do filtro, costuma-se preencher as bordas não calculadas com uma constante ou com o valor do ponto mais próximo. Este último procedimento é chamado de repetição de bordas. Outras abordagens podem ser seguidas, como definir os operadores de forma que tratem todos os casos especiais, assumir o quadro como sendo ciclicamente fechado ou atribuir o valor zero aos pontos fora dele.

3.2.2 Definição de Vizinhança

Um *pixel* qualquer do quadro é vizinho dele mesmo. A vizinhança é definida pela ordem que é definida pela distância do *pixel* de referência para o outro nas direções verticais e horizontais (SCHWARTZ, 1995).

O *pixel* localizado na posição (x, y) tem como vizinhos as variáveis com coordenadas (j, l) , tal que

$$0 < (j - x)^2 + (l - y)^2 \leq c, \quad (3.1)$$

em que c denota a ordem da vizinhança.

Conforme a Figura 3.5 e a Equação 3.1, pode-se definir o tipo de vizinhança como sendo de primeira, segunda, terceira e quarta ordem. As letras da Figura 3.5 representam a posição que se encontram em relação a posição central "s".

3.2.3 Tipos de Filtros

As técnicas de filtragem geralmente incorporam características do ruído conhecido *a priori* sobre bordas e propriedades do sistema visual humano para obter o efeito desejado das

	o_1	m	q_1	
o_2	v	u	z	q_2
l	t	s	t'	l'
q'_1	z'	u'	v'	o'_1
	q'_2	m'	o'_2	

Figura 3.5 Representação da vizinhança do pixel central com valor x_s .

técnicas de suavização, que têm como objetivos principais a remoção de ruído e a uniformização dos níveis de cinza dos *pixels* no quadro.

A filtragem no domínio espacial consiste na realização de operações diretamente nos *pixels* da imagem (GONZALEZ; WOODS, 2001), representado pela equação a seguir:

$$g(x,y) = T[f(x,y)] \quad (3.2)$$

na qual $f(x,y)$ é o valor do *pixel* cuja coordenada espacial é (x,y) e T é um operador sobre $f(x,y)$.

Neste tipo de filtragem o valor de cada *pixel* da imagem processada, $g(x,y)$, é obtido por meio de operações matemáticas realizadas diretamente sobre os *pixels* da imagem f . Para a obtenção do valor de cada *pixel* de g , o operador T pode ser aplicado a apenas um *pixel* de f ou a um conjunto de *pixels*, referido como janela.

Os filtros espaciais são implementados com máscaras de dimensões ímpares. Os tipos de filtros são passa baixas, passa altas, direcionais, passa banda.

3.2.4 Filtros Passa Baixas

Os filtros do tipo passa-baixa tendem a realçar as componentes de baixa frequência e atenuar as de alta frequência, ocorrendo perda de detalhes e redução do contraste do quadro, porém atenuam a influência de processos ruidosos provocados, por exemplo, por defeitos do sensor e erros na transformação matemática. Filtros passa-baixa suavizam os detalhes num quadro e reduzem o campo de níveis de cinza (contraste). Os filtros passa-faixa atenuam influências de ruídos periódicos (ruídos que são bem definidos), que são uma fonte de alta frequência.

Baixa frequência significa mudança gradual dos valores dos *pixels*. Ao aumentar o tamanho da janela na filtragem passa-baixa, os quadros resultantes ficam menos nítidos porque parte das componentes de alta frequência são removidas ou atenuadas (LABGIS, 2008).

Os filtros passa-baixa podem ser classificados como lineares e não-lineares. Os filtros passa-baixa lineares geram um quadro de saída mais homogêneo, ou mais suavizado, em

comparação com o quadro original. Em relação aos filtros lineares, os filtros passa-baixa não lineares possuem a vantagem de não afetar fortemente as áreas sem ruído homogeneizando o quadro. Como exemplo desse tipo de filtro há o filtro da mediana e o da moda.

Como os filtros da moda e da mediana são não lineares e, não existe uma formulação matemática linear que descreva o processo da sua aplicação sobre um quadro ou uma imagem. Esses filtros são definidos por sequências lógicas, ou algoritmos, que descrevem como as máscaras espaciais são aplicadas sobre o quadro ou imagem.

Filtro da Média

O filtro da média é um dos filtros mais simples que pode ser definido no domínio espacial. O valor da função $g(x,y)$, que representa o quadro filtrado, é definido pela média aritmética de todos os pontos superpostos à janela ou máscara. O objetivo desse filtro é a atenuação do ruído de alta frequência e suavização espacial dos quadros pelo do cálculo da média, pois cada ponto no quadro resultante é função de todos os pontos dentro de uma janela.

Apesar de conseguir atenuar o ruído, apresenta bons resultados apenas em regiões homogêneas, devido ao efeito de degradação de bordas. Esse filtro pode ser implementado com janelas de formato e tamanhos diversos, embora sejam mais comumente utilizadas as janelas quadradas de dimensões ímpares.

No filtro da média, o *pixel* central é calculado pela média aritmética dos *pixels* dentro da área da janela. Para cada janela tem-se as seguintes equações:

- Média com janela 2×2

$$g(x,y) = \frac{1}{N} \sum_{s=0}^m \sum_{t=0}^n f(x+s,y+t), \quad \forall W, \quad (3.3)$$

em que N é o número total de *pixels*, W é a janela que esta sendo usada e $m = n = (w - 1)$, na qual w é a dimensão da janela.

- Média com janela 3×3

$$g(x,y) = \frac{1}{N} \sum_{s=-m}^m \sum_{t=-n}^n f(x+s,y+t), \quad \forall W, \quad (3.4)$$

em que N é o número total de *pixels*, W é a janela que esta sendo usada e $m = n = (w - 2)$, na qual w é a dimensão da janela.

- Média com janela 4×4

$$g(x,y) = \frac{1}{N} \sum_{s=-m}^{m+1} \sum_{t=-n}^{n+1} f(x+s,y+t), \quad \forall W, \quad (3.5)$$

em que N é o número total de *pixels*, W é a janela que esta sendo usada e $m = n = (w - 2)/2$, na qual w é a dimensão da janela.

A filtragem espacial convencional utiliza um processo de geração do valor médio do quadro para cada janela. Assim, todas as regiões do quadro da janela W são igualmente afetadas. A média da filtragem espacial não leva em conta o efeito da diferença de níveis de cinza entre o *pixel* central e os *pixels* vizinhos e não tem sempre em conta a diminuição da influência dos *pixels* que se situem a uma maior distância do *pixel* central (ACHARYA; RAY, 2005).

A média é a medida de tendência central mais usada e tem as seguintes propriedades que justificam a sua utilização (STEVERSON, 1986) :

- A média pode sempre ser calculada;
- Para um dado conjunto de números, a média é única;
- A média é sensível a todos os valores do conjunto. Assim, se um valor se modifica, a média também se modifica;
- Adicionando (ou multiplicando) um peso a cada elemento do conjunto resulta em um média somada (ou multiplicada) por esse peso.

Filtro da Média Ponderada

Uma média ponderada é dada pela média de um conjunto de dados cujas entradas têm pesos variáveis. Cada peso depende da ordem de vizinhança do *pixel* a ser observado em relação ao *pixel* central. Nesse caso, a suavização é menos intensa porque o *pixel* central tem importância maior que os outros *pixels* da janela.

A filtragem linear com esse filtro de um quadro f de tamanho $M \times N$ por uma janela W de dimensão $(w \times w)$ é dada por (GONZALEZ; WOODS, 2001)

$$g(x, y) = \sum_{s=-m}^{m+1} \sum_{t=-n}^{n+1} \alpha(s, t) f(x + s, y + t), \quad (3.6)$$

com $\alpha(s, t)$ sendo o peso dado ao *pixel*.

Para a média ponderada são sugeridos nesse trabalho três formas, e cada forma tem uma relação com a ordem de vizinhança dos *pixels*:

1. A média ponderada considerando os *pixels* com vizinhança de primeira ordem é dada por

$$g(x, y) = \frac{1}{2}(x_s + \frac{1}{4}(x_{t'} + x_t + x_{u'} + x_u)). \quad (3.7)$$

2. A média ponderada considerando os *pixels* com vizinhança de segunda ordem é dada por

$$g(x, y) = \frac{1}{2} \left\{ x_s + \frac{1}{5} [x_{t'} + x_t + x_{u'} + x_u + \frac{1}{4}(x_{v'} + x_v + x_{z'} + x_z)] \right\}. \quad (3.8)$$

3. A média ponderada considerando os *pixels* com vizinhança de segunda ordem de uma segunda maneira, dada por

$$g(x,y) = \frac{1}{32} \{8x_s + 4(x_t + x_l + x_u + x_u) + 2(x_v + x_v + x_z + x_z)\}. \quad (3.9)$$

Filtro da Ordem

O filtro da ordem, operando em quadros, associa o K -ésimo valor dos níveis de cinza de uma janela $w \times w$, com $W = w^2$ elementos, arranjos de acordo com seu valor em ordem crescente, ao *pixel* central da janela.

O filtro da ordem é definido como tendo $p(1), p(2), \dots, p(W)$ os níveis de cinza dos W elementos de uma janela $w \times w$ do quadro, e $z(1), z(2), \dots, z(W)$ os mesmos níveis de cinza arranjos em ordem crescente (ARAÚJO, 1987; FACON, 2002). Assim $g(x,y)$ é definido por

$$g(x,y) = z(K), \quad (3.10)$$

K significando a ordem do filtro.

Os casos especiais $K = 1$ e $K = W$, conhecidos como os operadores min e max, respectivamente, são generalizações de operadores de expansão e compressão para aplicação em quadros ou imagens. O caso $K = (W + 1)/2$ é conhecido como o filtro da mediana.

Resultados apresentados por ARAÚJO (1987), Facon (2002) mostram que os filtros de ordem maior que o filtro da mediana alargam regiões claras do quadro proporcionalmente a sua ordem. Assim, objetos mais claros e pequenos, que são vistos com dificuldade no quadro original, são mostrados claramente pelo filtro max, que aumenta a área. Similarmente, os filtros de ordem menor que o filtro da mediana alargam regiões escuras e pequenas.

Filtro da Mediana

O filtro da mediana, ao contrário do filtro da média tem como vantagem manter os principais detalhes do quadro. A desvantagem é que o seu algoritmo é mais complexo, visto que utiliza ordenamento de valores.

O funcionamento do filtro da mediana envolve a verificação dos valores dos *pixels* vizinhos, em comparação com o *pixel* a ser calculado. Essa verificação consiste em ordenar os valores vizinhos de forma crescente e escolher a mediana dessa ordenação como o novo valor do *pixel* considerado (WU; RAO, 2006; YIN *et al.*, 1996), como mostrado em

$$g(x,y) = z((W + 1)/2). \quad (3.11)$$

Esse filtro suaviza o quadro, preservando a informação de bordas no quadro. A principal desvantagem destes filtros é a introdução de distorções verticais e de *aliasing*¹.

¹*Aliasing* ocorre quando existe a sobreposição do espectro.

Seu princípio de funcionamento baseia-se no fato de que elementos não pertencentes ao quadro tendem a se localizar nos extremos do conjunto ordenado. Isso ocorre pela expectativa de que os níveis de cinza dos pontos dentro de uma janela do quadro, de dimensões relativamente pequenas, apresentem um certo grau de similaridade. Como os pontos ruidosos dificilmente serão retornados como o valor da mediana, o filtro tem um bom desempenho na eliminação de ruídos, embora em alguns casos possa degradar os objetos da janela que tem uma variação rápida (AHMAD; SUNDARARAJAN, 1987).

Resultados apresentados em (ARAÚJO, 1987; FACON, 2002) indicam que o filtro da mediana é mais efetivo na redução de ruído impulsivo do que de ruído branco. Em relação ao filtro da média, o filtro da mediana se mostra melhor por ser relativamente insensível aos valores mais distantes do *pixel* central da janela, esses valores mais distantes do *pixel* central geralmente geram erros no filtro da média (STEVERSON, 1986).

Filtro da Moda

No filtro da moda o *pixel* central de uma região é substituído pelo valor mais frequente em sua vizinhança ou janela. A implementação desse filtro consiste no cálculo de um histograma dos valores vizinhos de um *pixel*. Em seguida, tomando aquele valor mais frequente como novo valor do *pixel* central considerado, ou seja, o valor do *pixel* central é dado por

$$g(x,y) = z_m \quad (3.12)$$

em que z_m é a intensidade de maior frequência na vizinhança W em torno de $g(x,y)$.

Quando não se encontra um único valor de moda pode-se utilizar o valor do *pixel* central ou, na falta deste, a moda mais próxima do valor central.

O filtro da moda é usado para homogeneizar quadros temáticos, ou para reduzir ruídos mantendo o máximo de informação no quadro. Esse filtro garante que o conjunto de valores digitais do quadro de saída é um subconjunto do domínio de valores do quadro de entrada. Assim, não são criados níveis digitais diferentes daqueles presentes no quadro de entrada.

Filtro Sigma

O filtro Sigma utiliza a distribuição estatística dos contadores digitais de uma janela móvel para estimar o valor do *pixel* central da janela, dito *pixel* de interesse (GONÇALVES; GARCIA, 1996). O filtro sigma (ARAÚJO, 1987), (LUKAC *et al.*, 2003) é motivado pela probabilidade da distribuição gaussiana. A suavização do ruído no quadro é alcançada pelo cálculo da média com apenas aqueles elementos que têm seus níveis de cinza ou intensidades dentro de uma faixa σ de intensidade, fixada pelo *pixel* central.

A distribuição do ruído em quadros é assumida gaussiana e a probabilidade dos 2σ é definida como sendo a probabilidade de uma variável aleatória estar dentro de uma faixa de dois desvios padrões do seu valor médio.

Sejam $p(i)$ a intensidade, ou nível de cinza, do i -ésimo *pixel* em uma janela $w \times w$ e σ o desvio padrão, dado pela Equação 3.13. O μ_f é a média dos valores de $p(i)$.

$$\sigma = \sqrt{\frac{1}{N} \sum (p(i) - \mu_f)^2} \quad (3.13)$$

Pode-se descrever o procedimento para se calcular o filtro sigma por:

- a) Estabeleça uma faixa de intensidade $(p(i) - T, p(i) + T)$, em que $T = 2\sigma$ (lembrando que: $p(i) - T$, tem como limite o valor 0 e $p(i) + T$, tem como limite o valor 255);
- b) Some todos os *pixels* da janela $w \times w$ que têm seus níveis de cinza dentro da faixa de intensidade;
- c) Calcule a média, dividindo a soma pelo número de *pixels*;
- d) então, $g(x, y) = \text{média}$.

$$g(x, y) = \frac{1}{N_T} \sum f_T(x, y) \quad (3.14)$$

em que $f_T(x, y)$ é o quadro limitado pela faixa de intensidade usada.

A faixa de intensidade dos dois-sigmas é geralmente grande o suficiente para incluir 95,5% dos *pixels* da mesma distribuição na janela, e ainda, na maioria das vezes, pequena o necessário para excluir *pixels* que representam bordas de contraste alto e pequeno.

3.3 Sistema Proposto

Nesta seção são descritos as etapas e os algoritmos dos sistemas de transcodificação propostos no trabalho, em que se faz a transcodificação de vídeos no domínio dos *pixels* com resolução CIF (352×288) para vídeos QCIF (176×144). O transcodificador foi feito usando a linguagem de programação C.

Para a redução da resolução espacial foi aplicado o subamostrador (S) com valor 2 ($352/176 = 288/144 = 2$) e os filtros da média, da média ponderada, da mediana, da moda e o sigma. Para os filtros da média, mediana, moda e sigma foram usadas janelas deslizantes com três dimensões ($w = 2, 3$ e 4).

Neste trabalho, também foi utilizada a janela $w = 1$ para o filtro da média. Essa janela contém apenas um *pixel* e pela observação da Fórmula 3.3 conclui-se que são eliminados os *pixels* das linhas e colunas ímpares do vídeo CIF, que resultou em uma eliminação simples.

Com o uso das três dimensões é importante analisar se há ou não sobreposição das janelas. Essas sobreposições ocorrem quando a janela é maior do que o fator do subamostrador (S), como pode ser mostrado nas Figuras 3.6 e 3.7. A sobreposição faz com que o *pixel* do vídeo transcodificado tenha uma maior correlação com os seus *pixels* vizinhos, o que produz

uma melhor codificação desse vídeo, mas gera também uma perda na qualidade do vídeo, já que o *pixel* do vídeo transcodificado foi produzido com *pixels* que tem pouca ou nenhuma dependência com o *pixel* central.

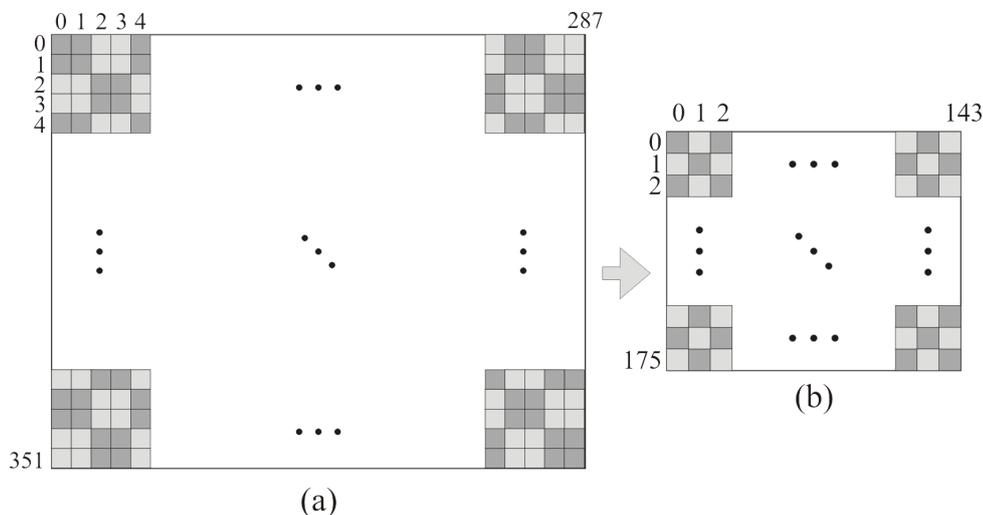


Figura 3.6 Transcodificação espacial com janela 2×2 . (a) Quadro CIF 352×288 , (b) Quadro QCIF 176×144

Na Figura 3.7 é ilustrado o caso $w = 3$, em que o *pixel* (1,1) da Figura 3.7(b) é obtido utilizando os *pixels* (1,1) a (1,3), (2,1) a (2,3) e (3,1) a (3,3) da Figura 3.7(a). Para se obter o *pixel* (1,2) da Figura 3.7(b) se utiliza a janela (1,3) a (1,5), (2,3) a (2,5) e (3,3) a (3,5). Da mesma forma, para se obter o *pixel* (2,1) da Figura 3.7(b), utiliza-se a janela da Figura 3.7(a) (3,1) a (3,3), (4,1) a (4,3) e (5,1) a (5,3).

Para a avaliação dos vídeos transcodificados métodos objetivos e subjetivos são usados. Na Figura 3.8 é apresentado o processo de avaliação utilizado neste trabalho, que consta da comparação dos vídeos transcodificados com o vídeo original por métrica objetiva e avaliação dos vídeos transcodificados com métricas subjetivas. Os vídeos originais utilizados neste trabalho foram os chamados vídeos de referências no formato QCIF. Os possíveis problemas referentes a esses vídeos são gerados por meio de não se saber a forma pela qual esses foram formados.

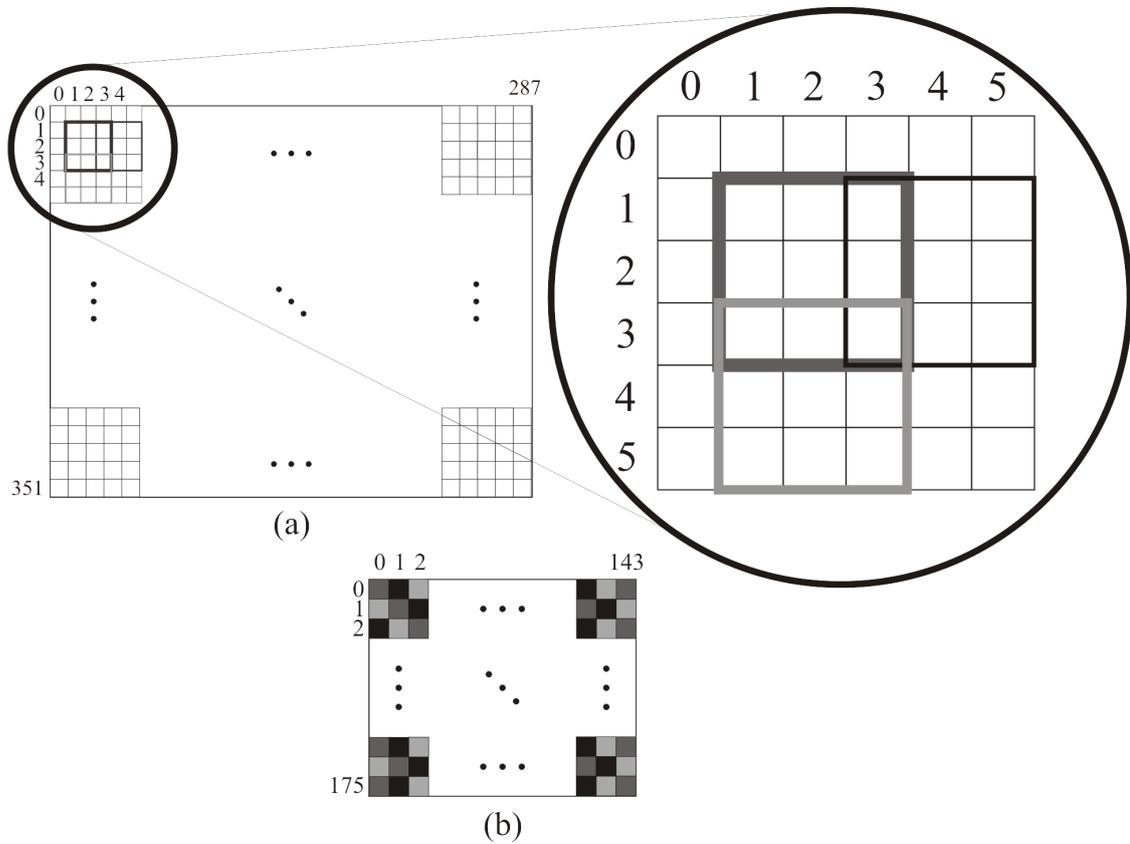


Figura 3.7 Técnica da média de *pixels* com janela 3×3 . (a) Quadro CIF 352×288 , (b) Quadro QCIF 176×144

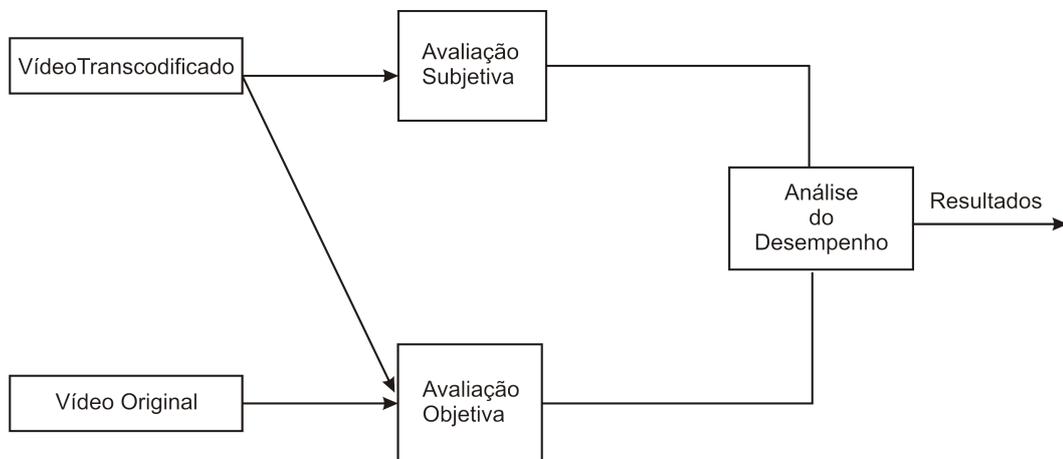


Figura 3.8 Processo de comparação do desempenho de vídeos transcodificados.

CAPÍTULO 4

Métodos de Avaliação

A importância da medida da qualidade é notada quando se leva em conta a diversidade de serviços disponíveis utilizando vídeo. A princípio, as pessoas pedem qualidade máxima. Mas há uma relação entre qualidade, preço, disponibilidade, acessibilidade, que cria um conceito de melhor qualidade dentro de certas condições. Os usuários de alguns tipos de serviço aceitam o sacrifício na parte da qualidade para redução de custos ou para que seja possível assistir o vídeo em um equipamento portátil (ARTHUR, 2002). Portanto, precisa-se medir ou qualificar um sinal dentro de certas condições para se determinar o quanto ele atende às expectativas.

A qualidade do vídeo pode ser medida utilizando métricas de qualidade objetiva ou de forma subjetiva por meio de experimentos com observadores humanos. Os procedimentos utilizados em um experimento da qualidade subjetiva de vídeo são descritos na recomendação BT.500 do ITU-T e a BT.500-11 do ITU-R para serviços de TV e a P.910 da ITU-T para aplicações multimídia (ITU-T, 1999). Mais detalhes desses métodos são encontrados no Apêndice A.

As medidas objetivas como SNR (*Signal-to-Noise Ratio*), PSNR (*Peak Signal to Noise Ratio*) ou MSE (*Mean Squared Error*) fazem medições em tempo real, comparando à imagem sob teste como a imagem original. Essas medidas mostram uma variação em relação à qualidade segundo observadores humanos, por consistirem de uma comparação matemática entre o sinal original e o sinal processado. Desta forma, um quadro deslocado de um *pixel* tem um PSNR baixo, mas um observador não nota a diferença entre a imagem deslocada e a original.

Por outro lado, as medidas subjetivas são realizadas por um grupo de pessoas, que subjetivamente classificam a qualidade do sinal processado de acordo com uma escala pré-determinada, comparando-o com outro sinal (BERNADINO, 1998), (KIM *et al.*, 2008).

4.1 Medidas Objetivas

As medidas de avaliação objetiva são mais rápidas e de menor custo do que as soluções subjetivas. Elas mostram degradações ainda imperceptíveis, e são fundamentais na organização de padrões. O conhecimento da existência da distorção imperceptível é importante para o

isolamento de sistemas defeituosos e na formação do conceito de limiares, a partir de que um determinado defeito passa a ser visível, e o quão longe o sistema em teste está deste limiar.

Pelas suas características de imparcialidade, confiabilidade, possibilidade de repetição e de baixo custo, métodos objetivos aumentam a competição entre provedores de serviço e equipamentos, e facilitam a especificação e avaliação de novos sistemas. A realização de medidas objetivas em tempo real abre a perspectiva de monitoração contínua da qualidade do vídeo (ARTHUR, 2002).

A avaliação objetiva de sinais de vídeo pode ser classificada em três categorias: (1) completamente referenciada, conhecida como FR ou *Full Reference*, quando ambos os sinais, original e processado, estão disponíveis para avaliação; (2) parcialmente referenciada, conhecida como RR ou *Reduced Reference*, quando apenas algumas amostras ou certas características do sinal original estão disponíveis; e (3) não referenciada, também conhecida como NR ou *No Reference*, quando apenas o sinal processado está disponível (FONSECA, 2008).

Para avaliação dos sistemas, é comum o uso do PSNR, ou do WSNR (*Weighted Signal to Noise Ratio*). A PSNR é definida pela Fórmula 4.1 para vídeos codificados com 8 bits.

$$\text{PSNR} = 10 \log_{10} \left[\frac{255^2}{\text{MSE}} \right] \quad (4.1)$$

e o MSE é dado pelo valor médio dos erros quadráticos entre os *pixels* do quadro original e do quadro decodificado. O MSE é dado pela fórmula 4.2:

$$\text{MSE} = \frac{1}{P} \sum_k \sum_{x,y} (f(x,y,k) - h(x,y,k))^2, \quad (4.2)$$

sendo P o número total de *pixels*, x e y os índices das linhas e colunas, respectivamente, k o número de quadros, f e h representando o quadro original e o modificado, respectivamente.

O PSNR é uma estimativa de qualidade da imagem reconstruída quando comparada com a original, de baixa complexidade. Calcula-se o erro médio quadrático (MSE) da imagem reconstruída, observando o nível de luminância de cada pixel. O uso do PSNR é bastante comum em função da facilidade de obtenção deste número. O MSE é bastante popular, apresenta expressões matemáticas simples, facilitando a manipulação analítica. Mas testes mostram que há pouca correlação de medidas baseadas no MSE com as observações subjetivas (WANG *et al.*, 2004).

Nas três últimas décadas têm sido estudadas maneiras de medir a qualidade de vídeo e imagens objetivamente de forma que essas métricas levem em consideração as características do sistema visual humano (HVS – *Human Visual System*). Em (MUNTEAN *et al.*, 2005) são apresentadas diversas outras métricas, dentre elas pode-se citar a métrica de similaridade estrutural (SSIM - *Structural Similarity*).

O SSIM é um algoritmo que utiliza a estatística da imagem para a avaliação de qualidade de imagens usando janelas. Os índices SSIM são calculados pelas janelas deslizantes, que se movem *pixel-a-pixel* do topo esquerdo até o canto inferior direito do quadro.

O SSIM considera a informação estrutural da imagem, ou seja, aqueles atributos que se refletem na estrutura dos objetos da cena, que dependem da média da luminância e do contraste da imagem. Em (WANG; BOVIK, 2002) é proposta uma indexação da imagem, na qual f e h são dois sinais não negativos que são alinhados um com o outro, e existem μ_f , μ_h , σ_f^2 , σ_h^2 e σ_{fh} . A média e o desvio padrão são considerados estimativas aproximadas da luminância e do contraste do sinal. A covariância (σ_{fh}) pode ser pensada como a medida de quanto um sinal é diferente do outro (WANG; BOVIK, 2004), (VRANJES *et al.*, 2007). Na Figura 4.1 é apresentado um diagrama simplificado para medição do índice de similaridade entre duas imagens f e h de acordo com (FONSECA, 2008).

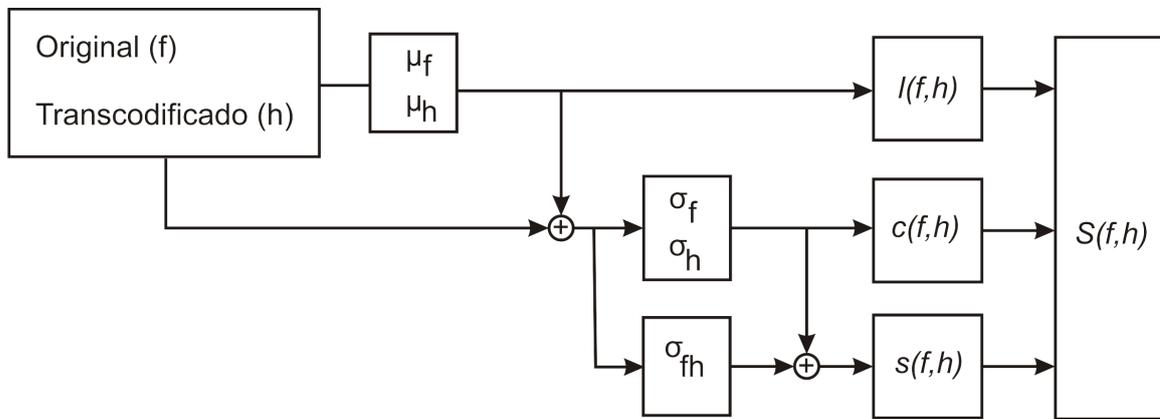


Figura 4.1 Diagrama simplificado para medição da similaridade estrutural entre duas imagens.

Se os dois sinais são representados por $f = \{f_i | i = 1, 2, \dots, P\}$ e $h = \{h_i | i = 1, 2, \dots, P\}$, as características estatísticas são dadas por:

$$\mu_f = \frac{1}{P} \sum_{i=1}^P f_i, \quad \mu_h = \frac{1}{P} \sum_{i=1}^P h_i, \quad (4.3)$$

$$\sigma_f^2 = \frac{1}{P-1} \sum_{i=0}^P (f_i - \mu_f)^2, \quad \sigma_h^2 = \frac{1}{P-1} \sum_{i=0}^P (h_i - \mu_h)^2, \quad (4.4)$$

$$\sigma_{fh} = \frac{1}{P-1} \sum_{i=0}^P (f_i - \mu_f)(h_i - \mu_h). \quad (4.5)$$

Em seguida deve-se remover a intensidade média de cada imagem e compará-las, obtendo-se a comparação de luminância $l(f, g)$, conforme mostrado na equação a seguir

$$l(f, h) = \frac{2\mu_f\mu_h}{\mu_f^2 + \mu_h^2}. \quad (4.6)$$

A comparação de contraste $c(f, g)$ é calculada usando a equação a seguir

$$c(f, h) = \frac{2\sigma_f\sigma_h}{\sigma_f^2 + \sigma_h^2}. \quad (4.7)$$

Após a subtração de luminância e a normalização de contraste, é feita uma comparação estrutural. Esta comparação é o coeficiente de correlação entre as imagens

$$s(f, h) = \frac{\sigma_{fh}}{\sigma_f\sigma_h}. \quad (4.8)$$

O $s(f, h)$ corresponde ao cosseno do ângulo entre os vetores $(f - \mu_f)$ e $(h - \mu_h)$, independente do comprimento desse vetor. Quando $(\mu_f^2 + \mu_h^2)(\sigma_f^2 + \sigma_h^2) \neq 0$, a similaridade indexada da medida entre f e h é dada em (WANG; BOVIK, 2002) e corresponde a

$$S(f, h) = l(f, h) c(f, h) s(f, h) = \frac{4\mu_f\mu_h\sigma_{fh}}{(\mu_f^2 + \mu_h^2)(\sigma_f^2 + \sigma_h^2)} \quad (4.9)$$

O problema com a Equação 4.9 é que quando $(\mu_f^2 + \mu_h^2)(\sigma_f^2 + \sigma_h^2)$ for igual a zero a medida resultante não está definida. Por esse motivo uma modificação da Equação 4.9 é feita (WANG; BOVIK, 2004), resultando na seguinte equação:

$$\text{SSIM}(f, h) = \frac{2(\mu_f\mu_h + C_1)(2\sigma_{fh} + C_2)}{(\mu_f^2 + \mu_h^2 + C_1)(\sigma_f^2 + \sigma_h^2 + C_2)} \quad (4.10)$$

as constantes C_1 e C_2 , são dadas por:

$$C_1 = (K_1B)^2, \quad C_2 = (K_2B)^2, \quad (4.11)$$

em que B é o valor dinâmico do *pixel* (para 8 *bits/pixel*, $B = 255$), K_1 e K_2 são duas constantes pequenas, dadas por $K_1 = 0,01$ e $K_2 = 0,03$, respectivamente, para este trabalho (WANG *et al.*, 2004).

Esta é a forma que o descritor SSIM se apresenta na literatura. Uma forma eficiente para implementação do SSIM é o janelamento, ou seja, subdividir as imagens sob comparação em blocos menores e utilizar as definições acima para obter o valor do SSIM para cada par de blocos, e então calcular o SSIM da imagem como um todo usando o valor médio. Para duas imagens subdivididas em D blocos cada, o índice pode ser calculado usando a seguinte expressão:

$$\text{SSIM}_i(f, h) = \frac{1}{D} \sum_{i=1}^D \text{SSIM}(f_i, h_i) \quad (4.12)$$

O SSIM tem as seguintes propriedades:

1. $\text{SSIM}(f, h) = \text{SSIM}(h, f)$;
2. $\text{SSIM}(f, h) \leq 1$;

3. SSIM (f, h) = 1, se e somente se $f = h$.

O ITU-T *Video Quality Expert Group* tem estudado extensivamente propostas de métricas objetivas para padronização e conclui dizendo que nenhuma delas atende a todas as necessidades em todas as condições, e, em consequência, atualmente não há solução objetiva capaz de substituir completamente os testes subjetivos.

4.2 Medidas Subjetivas

A avaliação subjetiva é a mais adequada para expressar os sentimentos e sensações dos observadores (REHME, 2007), e não é possível se caracterizar plenamente o desempenho dos sistemas com métricas objetivas, consequentemente, é necessário implementar as medidas subjetivas. Em (ITU-R, 2002) é dito que os métodos de avaliação subjetivos são usados para estabelecer o desempenho de sistemas de televisão usando medidas que mais diretamente antecipem as reações daqueles que verão os sistemas testados, métodos que podem ser encontrado no Apêndice A.

Porém, fatores externos interferem nos resultados das avaliações. O julgamento é afetado tanto por questões físicas como por puramente humanas. A cor e o nível de iluminação do ambiente onde se executa o ensaio, o tamanho e a distância da tela, mas também a composição dos avaliadores, seu estado físico, grau de cansaço, nível de instrução, ocupação profissional, humor, certamente individualizam parte da percepção do avaliador. O importante é se inserir padrões que deixem os resultados semelhantes quando oriundos de sistemas com características semelhantes, e que tais padrões reduzam a interferência dos fatores externos.

A recomendação do ITU-R (ITU-R, 2002) descreve uma série de critérios e metodologias definidas para testes mais críticos, realizados unicamente em laboratórios, e outros que permitem a avaliação da qualidade do ponto de vista do telespectador. Para ambos os casos há condições específicas e determinadas na recomendação sobre resolução, contraste, ângulo e distância do observador e tela, luminosidade e cor do ambiente. Como exemplo, a distância entre tela e observador deve satisfazer as regras de PVD (*Preferred Viewing Distance*), conforme Tabela 4.1. PVD é a melhor distância para visualização, e é dada em múltiplos das unidades de altura da tela, na qual o múltiplo (H) depende da altura da tela (REHME, 2007). Por exemplo, para um monitor de 20 polegadas, com a tela de altura de 30 cm, a melhor distância entre observador e a tela é o tamanho da altura da tela vezes e sete vezes a altura, ou seja, 2,1 m.

Há recomendações específicas para TV de alta definição, videoconferência e aplicações multimídia. Outros grupos também apresentaram algumas metodologias. Mas essencialmente possuem princípios similares e dada a abrangência e aceitação das recomendações ITU, considera-se suficiente o entendimento destas.

Estes experimentos envolvem pessoas consideradas de visão normal em ambientes controlados, seguindo padrões aceitos internacionalmente, como as recomendações ITU-R BT.500-

Tabela 4.1 PVD para imagens em movimento

Diagonal da tela (polegadas)		Altura da tela - H (cm)	PVD (H)
Tela relação 4/3	Tela relação 16/9		
12	15	18	9
15	18	23	8
20	24	30	7
29	36	45	6
60	73	91	5
> 100	> 120	> 153	3-4

11 (ITU-R, 2002) e a ITU-T P.910 (ITU-T, 1999) ambas da União Internacional de Telecomunicações. A avaliação subjetiva demanda sofisticados recursos, alto grau de habilidade e experiência de seus condutores, além de um longo período de tempo para sua conclusão. Os participantes devem ser, com diferentes idades, formações profissionais e que não sejam da área do projeto.

Para a validação deste tipo de avaliação é necessária a escolha das cenas. Um conjunto de cenas de vídeo de curta duração é escolhido. Estas cenas não devem apresentar distorções e devem representar trechos que caracterizem o contexto que está sendo avaliado. Cenas naturais e artificiais contendo cores fortes, texturas diversas, movimentos de câmera e de objetos em várias direções, contrastes suaves e fortes fazem parte destes conjuntos de cenas.

Na norma ITU-T P.910 (ITU-T, 1999), três métodos são recomendados para aplicações multimídia, sendo especificados os vídeos por classes e os vídeos usados nos testes sendo vídeos com codificação de vídeo PCM de 8 *bit/pixel*.

Entre as maneiras de apresentação das sequências de teste durante uma sessão do experimento pode-se destacar as seguintes (ITU-T, 1999):

- O método ACR (*Absolute Category Rating*) corresponde a uma metodologia na qual as sequências são apresentadas uma de cada vez de forma aleatória;
- O método DCR (*Degradation Category Rating*) corresponde a uma metodologia na qual as sequências são apresentadas em pares: o primeiro estímulo é sempre a sequência de referência ou original, enquanto o segundo estímulo apresentado corresponde a um dos sistemas sob teste;
- O método PC (*Pair Comparison*) corresponde a uma metodologia na qual as sequências de testes são apresentadas em pares, com a mesma sequência apresentada inicialmente para um sistema e, em seguida, para outro sistema.

Após análise dos três métodos foi escolhido o método PC, para que as sequências de teste sejam apresentadas em pares, fazendo uma melhor comparação entre os métodos de transcodificação.

O método PC consiste de sistemas sob ensaios (A, B, C, etc) que são organizados em todas as possíveis $n(n - 1)$ combinações do tipo AB, BA, CA, etc. Assim, todos os pares são exibidos em ambas as possíveis ordens (por exemplo, AB, BA). Após cada visualização uma sequência é preferida, no contexto do cenário do teste.

A sequência de apresentação utilizada no método PC é apresentada na Figura 4.2. O período de votação deve ser inferior ou igual a 10 s, dependendo do mecanismo de votação utilizado.

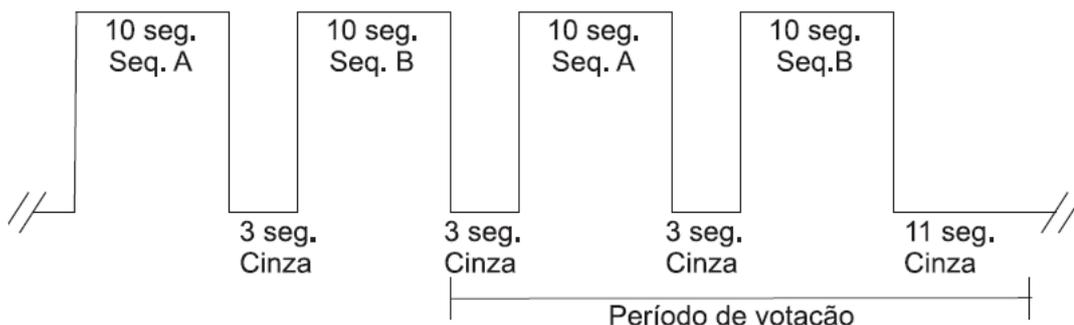


Figura 4.2 Apresentação dos estímulos no método PC.

Para o método PC, o número de repetições não precisa, em geral, ser considerado, pois o método implica na apresentação repetida das mesmas condições, embora em diferentes pares.

Para avaliação de amostras representativas usa-se a estatística, que é um conjunto de métodos especialmente apropriados à coleta, à apresentação (organização, resumo e descrição), à análise e à interpretação de dados de observação, tendo como objetivo a compreensão de uma realidade específica para a tomada da decisão.

4.3 Métricas Estatísticas

As opiniões a respeito das cenas originais e processadas resultam em um par de notas. A nota média e o desvio padrão é calculado para cada nota, resultando em uma variável denominada nota média na opinião dos observadores, *Mean Opinion Score* ou MOS. A média, o desvio padrão e o coeficiente de variação (CV) são medidas estatísticas. A média é dada por

$$\mu_{\alpha} = \frac{1}{A} \sum_{i=1}^A \alpha_i, \quad (4.13)$$

em que α_i são as amostras e A o número total de amostras. O desvio padrão, que é uma medida de dispersão, é dado por

$$\sigma = \sqrt{\frac{1}{A-1} \sum_{i=1}^A (\alpha_i - \mu_{\alpha})^2}. \quad (4.14)$$

O CV é uma média relativa à dispersão, útil para comparação e observação em termos relativos do grau de concentração em torno da média de séries distintas. É independente da unidade de medida utilizada, e é usada quando se deseja comparar a variação de conjuntos de observações que diferem na média ou são medidos em grandezas diferentes (unidades de medição diferentes). O CV é o desvio padrão expresso como uma porcentagem média.

$$CV = 100 \left(\frac{\sigma}{\mu_{\alpha}} \right) (\%) \quad (4.15)$$

Com o CV pode-se classificar a dispersão das séries, por (SHIGUTI; SHIGUTI, 2006):

- Dispersão baixa: $CV \leq 15\%$;
- Dispersão média: $15\% < CV < 30\%$;
- Dispersão alta: $CV \geq 30\%$.

4.3.1 Correlação Entre as Medidas

Objetivo do estudo correlacional é a determinação da força do relacionamento entre duas observações emparelhadas. A correlação indica até que ponto os valores de uma variável estão relacionados com os de outra. Para analisar a correlação é usado o coeficiente de correlação, dado por (STEVERSON, 1986):

$$r = \frac{\sum_{j=1}^A [(\alpha_j - \mu_{\alpha})(\beta_j - \mu_{\beta})]}{\sqrt{\sum_{j=1}^A (\alpha_j - \mu_{\alpha})^2} \sqrt{\sum_{j=1}^A (\beta_j - \mu_{\beta})^2}}, \quad (4.16)$$

na qual A é o número de amostras e α e β são as variáveis a serem relacionadas.

Os valores limites de r (coeficiente de correlação) são -1 e $+1$, isto é, o valor de r pertence ao intervalo $[-1,1]$. Assim (ROGER, 2007; SHMILDT, 2007):

- a) se não há correlação entre as variáveis, então $r = 0$.
- b) se a correlação é bem fraca entre as variáveis, então $0 < |r| < 0,2$;
- c) se a correlação é fraca entre as variáveis, então $0,2 \leq |r| < 0,4$;
- d) se a correlação é média entre as variáveis, então $0,4 \leq |r| < 0,7$;
- e) se a correlação é forte entre as variáveis, então $0,7 \leq |r| < 0,9$;
- f) se a correlação é muito forte entre as variáveis, então $0,9 \leq |r| < 1$;
- g) se a correlação entre duas variáveis é perfeita e positiva, então $r = 1$;
- h) se a correlação entre duas variáveis é perfeita e negativa, então $r = -1$.

CAPÍTULO 5

Resultados Experimentais

Neste capítulo são apresentados os resultados dos testes realizados para avaliar o desempenho do algoritmo de transcodificação espacial para TV móvel. Para estes testes foram utilizados dois dispositivos: um celular e um PDA. Os testes tiveram objetivo de estimar quais os parâmetros (taxa de *bits*, algoritmos, etc) apresentaram o melhor desempenho em termos de qualidade objetiva e subjetiva. Esses resultados foram obtidos usando métodos de aferição de qualidade objetivos e métricas subjetivas (experimentos psico-físicos) para os vídeos.

Para a realização dos testes foram utilizados dois dispositivos móveis, um celular NOKIA N95 e um PDA NOKIA N800. O N800 tem uma tela de 90×55 mm e resolução de Tela de 800×480 *pixels* com 65.536 cores. O N95 tem uma tela de 53×40 mm, 16 milhões de cores e 240×320 *pixels*.

O N800 foi escolhido devido às suas características e ao fato de se ter conseguido o empréstimo desse aparelho com o Instituto NOKIA de Tecnologia (INdT). O N95 foi escolhido por ser um dos aparelhos com as características mínimas necessárias para receber e reproduzir o sinal da TV móvel, tendo o melhor desempenho na época entre os celulares comercializados no Brasil.

Para a realização dos testes com esses dispositivos, foi preciso criar um conjunto de sequencias de teste. As sequencias de vídeos foram criadas utilizando o codificador H.264, sendo este o codificador recomendado para a TV móvel no Brasil (Apêndice B). Na codificação foi definida a taxa de quadros de 15 quadros por segundo, sendo esta a taxa de quadros mais comuns nas aplicações semelhantes (KIM *et al.*, 2008; KUMAR, 2007).

Após cada teste subjetivo foi entregue ao entrevistado um questionário com o objetivo de entender as expectativas do público em relação à TV móvel. O tipo de questionário utilizado é apresentado no Apêndice C.

5.1 Análise dos Vídeos em Diferentes Dispositivos

Para a avaliação da qualidade dos vídeos em diferentes dispositivos foram usados as sequencias de vídeo originais Paris, Mobile e Tempete (TRACE, 2008) (como mostrado na Figura 5.1), que são sequencias de vídeos padrões para testes na área de multimídia, transcodificados usando o filtro da média. Esses vídeos foram escolhidos para contemplar respectivamente, pouca, média e alta movimentação dos objetos na sequência de vídeo.



Figura 5.1 Vídeos Paris, Mobile e Tempete.

Os vídeos então foram transcodificados especialmente usando o filtro da média e para testar em aplicações semelhantes, esses foram codificados. Em seguida, mediu-se a PSNR do vídeo para avaliação dos vídeos transcodificados com diferentes janelas e o impacto no processo de codificação realizado pelo codificador.

Na Figura 5.2, são apresentadas três curvas da PSNR média (para todos os quadros) de cada vídeo, em função da dimensão das janelas consideradas no sistema proposto. Nessa figura percebe-se que o aumento do tamanho da janela deslizante utilizada provoca um aumento na PSNR média. Observe que um aumento da janela provoca uma maior redundância espacial nos quadros, o que aumenta significativamente a correlação entre *pixels* próximos dentro de um quadro, acarretando em uma melhor codificação, e portanto uma maior PSNR. Nessa figura também é possível perceber que com o aumento do movimento há uma diminuição na PSNR.

Na Figura 5.3, pode-se ver o fator de compressão obtido para cada vídeo, em que 100% implicaria em uma compressão total do vídeo. Observe, novamente, que com o aumento da janela obtém-se uma melhor compressão. Quanto menor o movimento da cena, maior será o fator de compressão obtido. Na janela 4×4 os impactos da textura e do movimento são quase imperceptíveis.

Uma característica dos vídeos transcodificados com as janelas 2×2 , 3×3 e 4×4 , é apresentar uma menor quantidade de componentes de alta frequência, as quais foram eliminadas pela operação do filtro da média. É interessante observar que para os vídeos transcodificados com as janelas deslizantes 3×3 e 4×4 existe a sobreposição das janelas, o que aumenta ainda mais a correlação entre *pixels*.

Também na Figura 5.2 é possível avaliar a influência da quantidade de movimentação apresentada por cada vídeo nos valores da PSNR. Conforme ilustrado, o vídeo Paris, com menor movimentação de objetos na cena, apresentou uma maior PSNR com relação aos demais vídeos

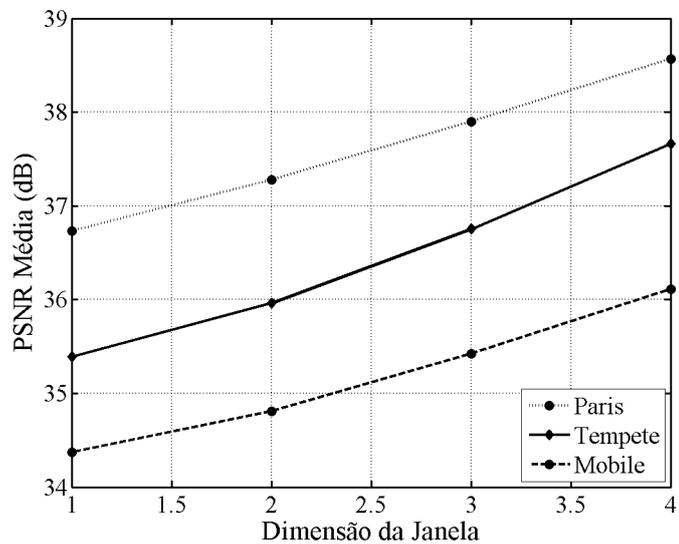


Figura 5.2 Gráfico da PSNR média versus dimensão da janela usando o filtro da média.

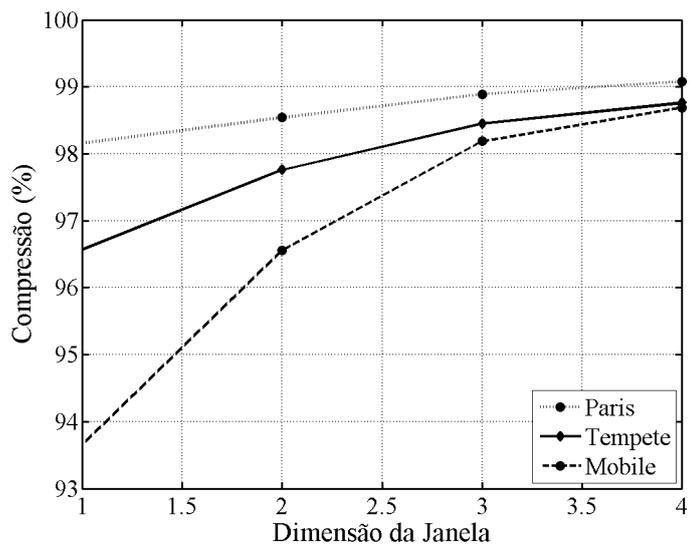


Figura 5.3 Compressão versus dimensão da janela usando o filtro da média.

devido à maior correlação entre os quadros. As curvas da PSNR apresentadas estão de acordo com os valores encontrados na literatura (AHMAD *et al.*, 2005).

Para a análise dos vídeos em diferentes dispositivos foi usada a medida subjetiva por meio do método PC. Para cada um dos dispositivos utilizados foram realizados testes com 20 participantes entre homens e mulheres com faixa de idade entre 15 e 55 anos, todos não especialistas da área de vídeo. Os testes duraram em média 30 minutos para o N800 e 35 minutos para o N95. Os participantes forneceram valores discretos de 0 a 10, para cada vídeo assistido, correspondendo ao seu julgamento da qualidade do vídeo. Cada participante assistiu seis vezes cada vídeo gerando assim 120 amostras por vídeo.

Na Tabela 5.1, utiliza-se como dispositivo o N800. Nela estão o valor médio das notas (MOS - *Mean Opinion Score*), o desvio padrão e o coeficiente de variação para os 12 vídeos. Para todos os experimentos, o nome do vídeo nas tabelas inclui o número referente à janela empregada.

Tabela 5.1 MOS, desvio padrão e coeficiente de variação dos testes subjetivos para o N800.

Vídeos	Nome	MOS	Desvio Padrão	CV %
1	Mobile 1	5,4750	0,1247	2,2770
2	Mobile 2	6,8333	0,1346	1,9696
3	Mobile 3	7,4786	0,1174	1,5704
4	Mobile 4	6,2750	0,1235	1,9687
5	Paris 1	5,7250	0,1307	2,2829
6	Paris 2	7,1500	0,1133	1,5846
7	Paris 3	6,9328	0,1111	1,6032
8	Paris 4	5,9417	0,1118	1,8822
9	Tempete 1	6,2542	0,1187	1,8972
10	Tempete 2	7,2083	0,1224	1,6974
11	Tempete 3	6,8917	0,1090	1,5813
12	Tempete 4	5,8833	0,1099	1,8674

Para os melhores resultados de cada vídeo o MOS ficou acima do valor 7, como pode ser observado na Figura 5.4, o que significa uma boa aceitação dos entrevistados. Para os vídeos Tempete e Paris os melhores resultados foram obtidos para a janela 2×2 , enquanto que para o Mobile a janela 3×3 obteve o melhor resultado. Em relação ao coeficiente de variação a janela 1×1 obteve os piores resultados o que mostra que ele teve uma maior variação, porém os resultados estão abaixo dos valores encontrados na literatura (FECHINE, 2000) e são ditos de dispersão baixa (SHIGUTI; SHIGUTI, 2006).

Na Tabela 5.2, são apresentados os resultados para o dispositivo N95. Na tabela são fornecidos o MOS, o desvio padrão e o coeficiente de variação para os 12 vídeos de testes.

Como pode ser visto nas Figuras 5.4 e 5.5, exceto para o caso do vídeo Mobile, a média dos vídeos transcodificados usando a janela 2×2 obteve valores mais altos que os demais obtidos utilizando janelas de tamanhos diferentes. Os vídeos transcodificados usando a janela 3×3

Vídeos	Nome
1	Mobile 1
2	Mobile 2
3	Mobile 3
4	Mobile 4
5	Paris 1
6	Paris 2
7	Paris 3
8	Paris 4
9	Tempete 1
10	Tempete 2
11	Tempete 3
12	Tempete 4

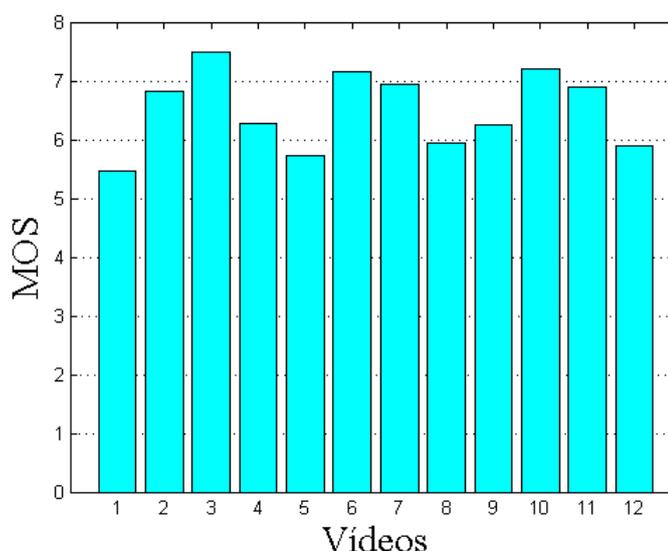


Figura 5.4 Valor médio dos valores dados para cada vídeo usando o N800.

Tabela 5.2 MOS, desvio padrão e coeficiente de variação dos testes subjetivos para o N95.

Vídeos	Nome	MOS	Desvio Padrão	CV %
1	Mobile 1	6,5093	0,1173	1,8020
2	Mobile 2	7,0278	0,1031	1,4675
3	Mobile 3	7,5741	0,1221	1,6127
4	Mobile 4	6,9352	0,1186	1,7101
5	Paris 1	6,5333	0,1058	1,6187
6	Paris 2	7,1759	0,1051	1,4647
7	Paris 3	7,1296	0,0994	1,3942
8	Paris 4	6,5370	0,1351	2,0664
9	Tempete 1	7,1759	0,1168	1,6277
10	Tempete 2	7,7500	0,1086	1,4011
11	Tempete 3	7,4537	0,1208	1,6213
12	Tempete 4	6,7685	0,1469	2,1696

obtiveram maiores valores devido ao vídeo Mobile ter mais textura. Para os maiores valores de cada vídeo, as notas foram superiores a 7.

A pior MOS foi obtida para o vídeo transcodificado usando a janela 1×1 , exceto para o vídeo Tempete que obteve como pior resultado o vídeo transcodificado usando a janela 4×4 , para o N800. Para o N95 a diferença ocorreu no vídeo Paris, que obteve MOS praticamente iguais para os vídeos transcodificados usando as janelas 1×1 e 4×4 . A pior MOS encontrada foi 5,4750 para o N800 e 6,5093 para o N95, ambos para o vídeo Mobile na janela 1×1 . Os melhores resultados foram obtidos para o N95 em relação ao N800. Vale salientar que o N95 tem uma tela menor, resultando em menor percepção das distorções pelos entrevistados.

Vídeos	Nome
1	Mobile 1
2	Mobile 2
3	Mobile 3
4	Mobile 4
5	Paris 1
6	Paris 2
7	Paris 3
8	Paris 4
9	Tempete 1
10	Tempete 2
11	Tempete 3
12	Tempete 4

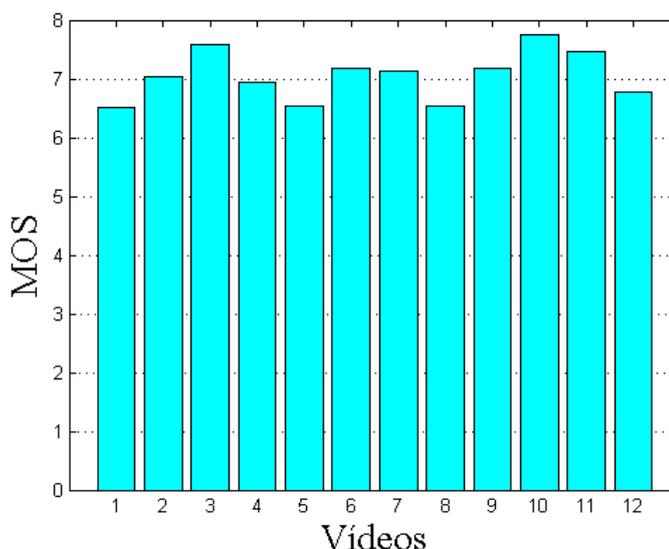


Figura 5.5 Valor médio dos valores (MOS) dados para cada vídeo usando o N95, com taxa de *bit* de 243 *kbit/s*.

Para todos os vídeos o desvio padrão obtido teve um valor considerado pequeno, variando entre 0,109 e 0,1346 para o N800 e, 0,0994 e 0,1469 para o N95. Observando as Figuras 5.4 e 5.5 pode se concluir que os resultados foram melhores usando o N95.

Em relação ao coeficiente de variação, é possível verificar que os resultados obtidos são ditos de dispersão baixa (SHIGUTI; SHIGUTI, 2006). O coeficiente de variação variou entre 1,3942 e 2,2829, tendo uma pequena variação entre o máximo e o mínimo. Os menores resultados estão de acordo com os melhores resultados da média.

5.2 Avaliação da Taxa de *Bit* de Vídeo

Para a avaliação das medidas subjetivas dos vídeos foram usados três vídeos originais no formato QCIF, o Walk, o Akiyo e o Coastguard (TRACE, 2008), como mostrado na Figura 5.2. Para os testes foi utilizado o dispositivo N95. Cada um deles foi codificado com três taxas de *bits* diferentes: 111 *kbit/s*, 243 *kbit/s* e 318 *kbit/s*. Estas taxas foram escolhidas por serem as mais comumente utilizadas em aplicações de TV móvel (KNOCHE; SASSE, 2006).



Figura 5.6 Vídeos Walk, Akiyo e Coastguard

Cada participante assistiu quatro vezes cada vídeo gerando 60 amostras por vídeo e cada teste durou em média 19,75 minutos. A análise dos vídeos foi avaliada pelo MOS dos participantes, pelo desvio padrão e pelo coeficiente de variação.

Na Tabela 5.3, é mostrado o MOS, o desvio padrão e o coeficiente de variação para os nove vídeos (cada vídeo original com três taxas de *bits* diferentes) usados nos testes. O nome do vídeo na Tabela inclui o nome do vídeo original e o número da taxa de *bit* que esse foi codificado.

Tabela 5.3 MOS, desvio padrão e o coeficiente de variação dos vídeo originais com diferentes taxas.

Vídeos	Nome	MOS	Desvio Padrão	CV %
1	Walk 111	5,9500	0,1822	3,0622
2	Walk 243	7,7848	0,1320	1,6954
3	Walk 318	7,8250	0,1150	1,4694
4	Akiyo 111	6,6500	0,1982	2,9800
5	Akiyo 243	7,1899	0,1766	2,4561
6	Akiyo 318	7,2750	0,1715	2,3578
7	Coastguard 111	6,5375	0,1831	2,8008
8	Coastguard 243	7,4000	0,1536	2,0763
9	Coastguard 318	7,5875	0,1304	1,7180

Os valores do coeficiente de variação, mostra que os resultados possuem uma alta confiabilidade. Entretanto comparando estes resultados com os resultados apresentados na Seção 5.1, podemos observar que houve um aumento nos valores do coeficiente de variação, já que variou de 1,4694 a 3,0622 e todos também podem ser ditos de dispersão baixa. Os valores dos coeficientes de variação diminuem à medida que a taxa de *bit* aumenta.

Na Figura 5.7, é apresentado um gráfico de barra do MOS (obtidas do experimento) para todas as sequencias de teste. Observe que os valores MOS obtidos para vídeos com taxa 243 *kbit/s* e 318 *kbit/s* apresentam resultados similares, enquanto que os vídeos com taxa de *bit* de 111 *kbit/s* obtiveram resultados piores. Os vídeos com taxa de *bit* de 318 *kbit/s*, obtiveram melhor avaliação subjetiva, mas a diferença não foi significativa entre essa taxa de *bit* e de 243 *kbit/s*.

Após esse teste foi feito um outro teste para comparar se há uma diferença em termos de qualidade subjetiva entre os vídeos codificados com taxas de 243 *kbit/s* e 318 *kbit/s* e transcodificados espacialmente utilizando o filtro da média, com quatro tamanhos de janelas diferentes. Foram utilizados os vídeos Paris, Mobile e Tempete. Os resultados coletados do experimento psico-físico são apresentado na Tabela 5.4 e na Figura 5.8.

Pode-se observar nas Figuras 5.5 e 5.8 que, exceto para o vídeo Mobile, os valores do MOS obtidas não apresentam diferença significativa. Para os vídeos codificados a uma taxa de 243 *kbit/s* os resultados obtidos foram um pouco melhores. Os valores do coeficiente

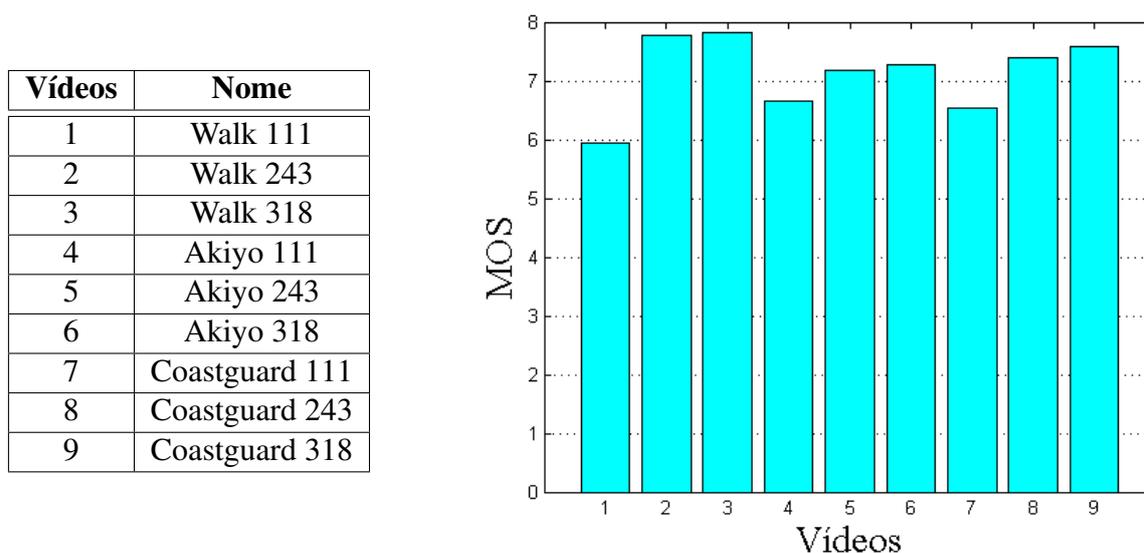


Figura 5.7 MOS obtidos para cada uma das sequencias de teste.

Tabela 5.4 Valores do MOS, desvio padrão e coeficiente de variação dos MOS obtidos dos testes subjetivos utilizando o N95 e taxa de *bit* de 318 kbit/s.

Vídeos	Nome	MOS	Desvio Padrão	CV %
1	Mobile 1	6.8917	0.1394	2.0231
2	Mobile 2	7.2583	0.1035	1.4260
3	Mobile 3	7.0250	0.1233	1.7545
4	Mobile 4	6.1261	0.1185	1.9340
5	Paris 1	5.8333	0.1499	2.5689
6	Paris 2	7.1453	0.1053	1.4739
7	Paris 3	6.6667	0.1260	1.8898
8	Paris 4	5.7083	0.1150	2.0143
9	Tempete 1	7.0083	0.1289	2.0578
10	Tempete 2	7.5630	0.1289	1.7042
11	Tempete 3	7.0000	0.1417	2.0245
12	Tempete 4	5.8250	0.1368	2.3489

de variação obtidos também não tiveram uma diferença significativa, variando entre 1,4260 e 2,5689.

5.3 Medidas Objetivas dos Vídeos Transcodificados

5.3.1 MSE e PSNR

Para avaliação do transcodificador foram gerados vídeos usando cada uma das técnicas propostas, para os vídeos originais Mobile, News e Foreman, como mostrado na Figura 5.9. Esses vídeos foram escolhidos por apresentarem áreas com uma boa quantidade de movimento

Vídeos	Nome
1	Mobile 1
2	Mobile 2
3	Mobile 3
4	Mobile 4
5	Paris 1
6	Paris 2
7	Paris 3
8	Paris 4
9	Tempete 1
10	Tempete 2
11	Tempete 3
12	Tempete 4

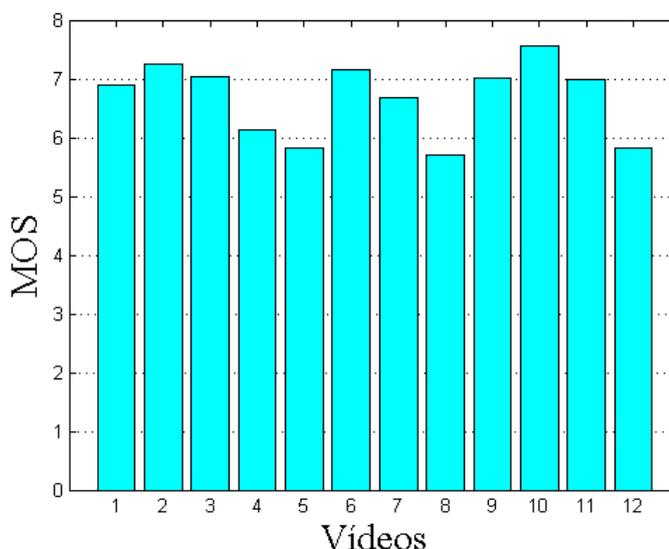


Figura 5.8 Valor do MOS para os vídeos com taxa de 318 kbit/s.

e textura. Um fator importante que foi levado em consideração é existir esses vídeos padrões nas resoluções CIF e QCIF. Essa condição é importante pelo fato de poder usar o vídeo na resolução QCIF como sendo o vídeo original para se realizar os testes objetivos.



Figura 5.9 Vídeo Mobile, News e Foreman

Os vídeos podem ser classificados pela quantidade de movimento e pela textura. Quanto a quantidade movimentação o vídeo News é o que tem a maior quantidade de movimentação, seguido pelo vídeo Foreman e depois o vídeo Mobile. Quanto textura o vídeo Mobile é o que tem maior textura, seguido pelo vídeo News e depois o vídeo Foreman.

O MSE é a primeira métrica utilizada para aferir o desempenho dos vídeos. A Tabela 5.5 mostra os resultados e a Figura 5.10 apresenta o gráfico do MSE.

Uma outra forma de avaliar a eficiência do transcodificador é pela PSNR. A Tabela 5.6 mostra os resultados e a Figura 5.11 apresenta o gráfico da PSNR.

Para o vídeo original Mobile as sequencias de teste que apresentaram os melhores resultados foram os processados com o filtro sigma 4×4 , o filtro sigma 2×2 e o filtro mediana 4×4 . Para o vídeo original News os melhores resultados correspondem aos vídeos processados com o filtro sigma 2×2 , o filtro mediana 2×2 e o filtro mediana 4×4 . Para o vídeo original Foreman

Tabela 5.5 Resultados do MSE.

Numeração	Filtros	Mobile	News	Foreman
1	Eliminação Simples	10,47	1,68	0,83
2	Média 2×2	7,83	3,69	2,51
3	Média 3×3	4,77	0,87	0,34
4	Média 4×4	5,53	1,96	0,92
5	Mediana 2×2	3,61	0,38	0,42
6	Mediana 3×3	6,18	1,18	0,46
7	Mediana 4×4	3,52	0,56	0,42
8	Moda 2×2	7,09	2,14	1,88
9	Moda 3×3	13,91	3,57	1,9
10	Moda 4×4	20,28	4,04	2,07
11	Média ponderada 1	6,87	1,1	0,43
12	Média ponderada 2	6,56	1,05	0,4
13	Média ponderada 3	5,28	0,88	0,32
14	Sigma 2×2	3,18	0,3	0,36
15	Sigma 3×3	5,02	0,89	0,36
16	Sigma 4×4	3,17	0,65	0,45

Tabela 5.6 Resultados do PSNR.

Numeração	Filtros	Mobile	News	Foreman
1	Eliminação Simples	37,93	45,87	48,69
2	Média 2×2	39,19	42,46	44,13
3	Média 3×3	41,35	48,76	52,79
4	Média 4×4	40,71	45,20	48,51
5	Mediana 2×2	42,55	52,39	51,94
6	Mediana 3×3	40,22	47,44	51,51
7	Mediana 4×4	42,67	50,66	51,93
8	Moda 2×2	39,62	44,82	45,38
9	Moda 3×3	36,70	42,61	45,34
10	Moda 4×4	35,06	42,06	44,97
11	Média ponderada 1	39,76	47,71	51,80
12	Média ponderada 2	39,96	47,93	52,07
13	Média ponderada 3	40,91	48,69	53,02
14	Sigma 2×2	43,1	53,33	52,62
15	Sigma 3×3	41,12	48,65	52,62
16	Sigma 4×4	43,12	50,03	51,59

Vídeos	Nome
1	Eliminação Simples
2	Média 2 × 2
3	Média 3 × 3
4	Média 4 × 4
5	Mediana 2 × 2
6	Mediana 3 × 3
7	Mediana 4 × 4
8	Moda 2 × 2
9	Moda 3 × 3
10	Moda 4 × 4
11	Média ponderada 1
12	Média ponderada 2
13	Média ponderada 3
14	Sigma 2 × 2
15	Sigma 3 × 3
16	Sigma 4 × 4

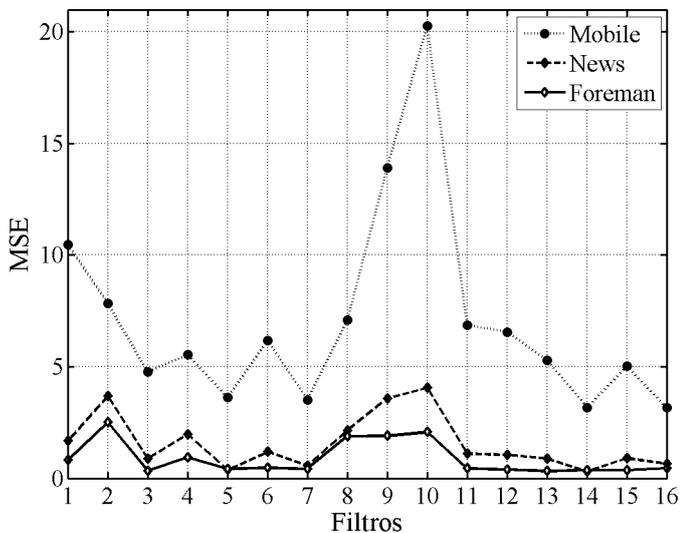


Figura 5.10 Gráfico da MSE para os vídeos transcodificados.

Vídeos	Nome
1	Eliminação Simples
2	Média 2 × 2
3	Média 3 × 3
4	Média 4 × 4
5	Mediana 2 × 2
6	Mediana 3 × 3
7	Mediana 4 × 4
8	Moda 2 × 2
9	Moda 3 × 3
10	Moda 4 × 4
11	Média ponderada 1
12	Média ponderada 2
13	Média ponderada 3
14	Sigma 2 × 2
15	Sigma 3 × 3
16	Sigma 4 × 4

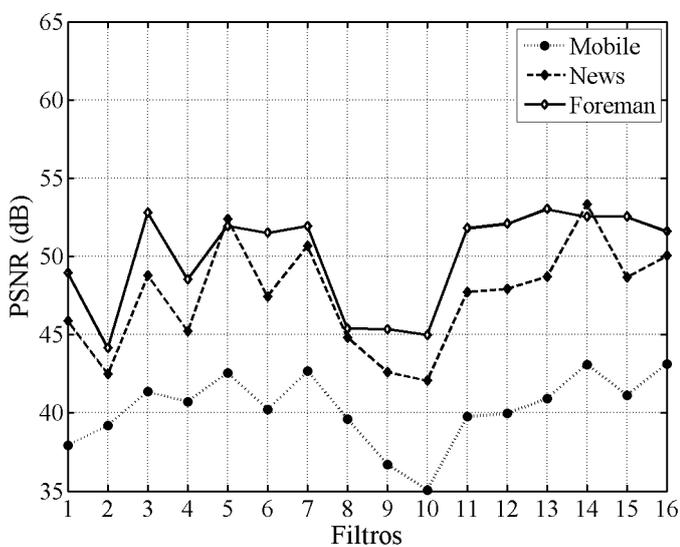


Figura 5.11 Gráfico da PSNR para os vídeos transcodificados.

os melhores resultados correspondem aos vídeos processados com o filtro média ponderada 3, o filtro média 3×3 e o filtro sigma 2×2. Estes resultados são apresentados na Figura 5.11.

5.3.2 SSIM

Para a avaliação dos vídeos transcodificados usando o método SSIM, como apresentado na seção 4.1, foram utilizados os mesmos vídeos originais utilizados na Subseção 5.3.1. Os valores da métrica SSIM para cada um dos vídeos processados são apresentados na Tabela 5.7.

Tabela 5.7 Resultados do método SSIM.

Numeração	Filtros	Mobile	News	Foreman
1	Eliminação Simples	0,9713	0,9793	0,9783
2	Média 2×2	0,9842	0,9618	0,9739
3	Média 3×3	0,9785	0,9744	0,9841
4	Média 4×4	0,9511	0,9560	0,9701
5	Mediana 2×2	0,9910	0,9901	0,9942
6	Mediana 3×3	0,9828	0,9805	0,9882
7	Mediana 4×4	0,9578	0,9704	0,9799
8	Moda 2×2	0,9765	0,9752	0,9764
9	Moda 3×3	0,9707	0,9705	0,9726
10	Moda 4×4	0,9647	0,9732	0,9725
11	Média ponderada 1	0,9871	0,9819	0,9887
12	Média ponderada 2	0,9877	0,9814	0,9882
13	Média ponderada 3	0,9859	0,9786	0,9867
14	Sigma 2×2	0,9918	0,9885	0,9939
15	Sigma 3×3	0,9813	0,9783	0,9860
16	Sigma 4×4	0,9593	0,9636	0,9751

Pode-se observar da Tabela 5.7 que os melhores resultados para o vídeo Mobile foram obtidos utilizando os filtros sigma 2×2 , mediana 2×2 e média ponderada 2. Para os vídeos News e Foreman os melhores resultados obtidos utilizando os filtros mediana 2×2 , sigma 2×2 e média ponderada 1. Estes resultados são apresentados na Figura 5.12.

Com os resultados dos métodos PSNR e SSIM foi possível encontrar a correlação entre as medidas. A correlação em relação ao vídeo Mobile obtido foi de 0,1408, que é uma correlação bem fraca. Para o vídeo News a correlação encontrada entre as medidas foi de 0,5424, que é uma correlação média. Para o vídeo Foreman a correlação entre as medidas obtida foi de 0,7492, que é uma correlação forte.

5.3.3 Tempo de Processamento

Em relação ao tempo de processamento é possível analisar um aumento no tempo à medida que se aumenta a janela do filtro utilizado, como mostrado na Tabela 5.8. Nessa Tabela também pode-se ver um aumento no tempo de processamento entre a média e a mediana e uma elevação entre a mediana e a moda.

O algoritmo de transcodificação que utilizou o filtro média ponderada obteve um tempo de execução semelhante ao algoritmo que utilizou o filtro média. Já o algoritmo de transcodificação que utilizou o filtro sigma obteve um tempo entre o tempo gasto pela transcodificação

Vídeos	Nome
1	Eliminação Simples
2	Média 2×2
3	Média 3×3
4	Média 4×4
5	Mediana 2×2
6	Mediana 3×3
7	Mediana 4×4
8	Moda 2×2
9	Moda 3×3
10	Moda 4×4
11	Média ponderada 1
12	Média ponderada 2
13	Média ponderada 3
14	Sigma 2×2
15	Sigma 3×3
16	Sigma 4×4

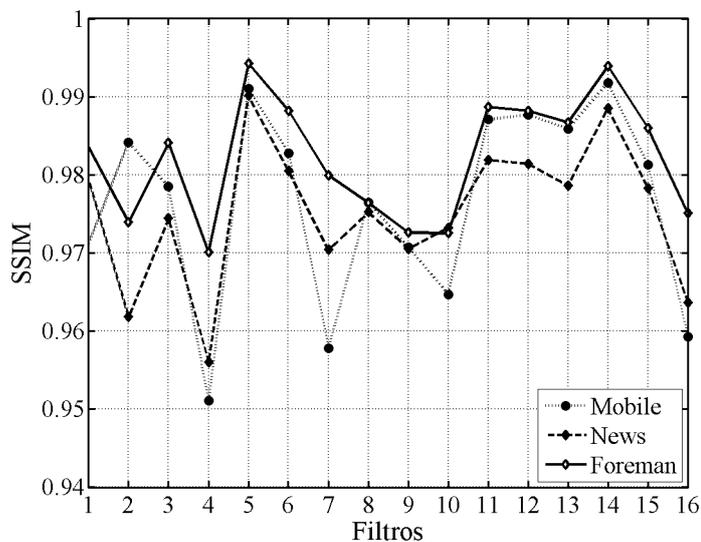


Figura 5.12 Gráfico do SSIM para os vídeos transcódificados.

especial utilizando os filtros mediana e a moda cujo o valor é apresentado na Tabela 5.8. O valor do tempo gasto utilizando o filtro sigma varia de acordo com a janela, por isso na Tabela 5.8 é apresentado um valor médio. Esses valores foram obtidos usando um computador que operava apenas o programa do transcodificador e que tem a seguinte configuração: Processador AMD Athlon(tm) 64 X2 Dual Core Processor 4400+, 231 GHz, 2,00 GB de RAM.

Tabela 5.8 Tempo de processamento de um vídeo

Metodo de Transcodificação	Tempo(segundos)
Eliminação Simples	0,47
Média 2×2	1,30
Média 3×3	1,13
Média 4×4	3,89
Mediana 2×2	1,59
Mediana 3×3	5,00
Mediana 4×4	13,69
Moda 2×2	7,78
Moda 3×3	7,22
Moda 4×4	56,47
Média ponderada 1	0,75
Média ponderada 2	1,19
Média ponderada 3	3,42
Sigma 2×2	5,76
Sigma 3×3	12,06
Sigma 4×4	20,50

5.3.4 Erro Provocado pela Codificação

Para a realização dos testes subjetivos é necessário que os vídeos transcodificados sejam codificados, para que eles reproduzam no dispositivo móvel. Assim, os vídeos foram codificados usando o codificador H.264 para que seguissem as normas da TV digital no Brasil. A codificação gera erros, e o codificador avalia os vídeos codificados usando a PSNR e o SSIM.

Os valores da PSNR obtidos são apresentados na Tabela 5.9 e na Figura 5.13. Para os vídeos originais codificados, a PSNR avaliada foi de 40 dB para o vídeo Mobile, 50,28 dB para o vídeo News e 48,69 dB para o vídeo Foreman.

Tabela 5.9 Resultados do PSNR dos vídeos codificados após a transcodificação

Numeração	Filtros	Mobile	News	Foreman
1	Eliminação Simples	31,47	49,79	42,07
2	Média 2×2	35,61	50,07	45,34
3	Média 3×3	41,16	50,71	48,95
4	Média 4×4	42,84	51,00	49,62
5	Mediana 2×2	34,97	50,05	45,18
6	Mediana 3×3	35,89	50,33	45,84
7	Mediana 4×4	39,13	50,70	47,61
8	Moda 2×2	31,36	48,22	41,29
9	Moda 3×3	30,38	48,46	40,29
10	Moda 4×4	29,12	43,17	39,22
11	Média ponderada 1	35,27	50,06	45,15
12	Média ponderada 2	35,66	50,10	45,44
13	Média ponderada 3	38,74	50,45	47,65
14	Sigma 2×2	36,08	49,82	45,64
15	Sigma 3×3	39,12	50,07	47,152
16	Sigma 4×4	41,06	50,3	48,24

Para o vídeo Mobile os vídeos gerados pelos filtros média 3×3 , média 4×4 e sigma 4×4 obtiveram resultados melhores do que os resultados do vídeo original. Para o vídeo News os vídeos gerados pelos filtros média 3×3 , média 4×4 , mediana 3×3 , mediana 4×4 , média ponderada 3 e sigma 4×4 obtiveram resultados melhores que os resultados do vídeo original. Para o vídeo Foreman os vídeos gerados pelos filtros média 3×3 e média 4×4 obtiveram resultados melhores que os resultados do vídeo original.

Comparando a Tabela 5.9 com a Tabela 5.6 é visto que para os vídeos Mobile e Foreman os erros na codificação são maiores que os erros provocados pela transcodificação, exceto para os vídeos gerados utilizando os filtros média 2×2 e média 4×4 para o vídeo Foreman e o filtro média 4×4 para o vídeo Mobile. Já para o News os erros na codificação são menores que os erros da transcodificação, exceto para os vídeos gerados utilizando os filtros mediana 2×2 e sigma 3×3 .

A correlação entre as medidas da PSNR dos vídeos transcodificados e dos vídeos transcodificados espacialmente após a codificação mostra que para os vídeos Foreman e News existe

uma correlação média, com 0,68 e 0,531, respectivamente. Para o vídeo Mobile a correlação entre as medidas é forte com 0,7468.

Observa-se na Figura 5.13 que os vídeos codificados e transcodificados utilizando o filtro da moda obtiveram os piores resultados. Destes resultados, foi possível observar que vídeos cujo conteúdo possui mais movimento produziram maiores valores de erro, por serem mais difíceis de codificar.

Vídeos	Nome
1	Eliminação Simples
2	Média 2×2
3	Média 3×3
4	Média 4×4
5	Mediana 2×2
6	Mediana 3×3
7	Mediana 4×4
8	Moda 2×2
9	Moda 3×3
10	Moda 4×4
11	Média ponderada 1
12	Média ponderada 2
13	Média ponderada 3
14	Sigma 2×2
15	Sigma 3×3
16	Sigma 4×4

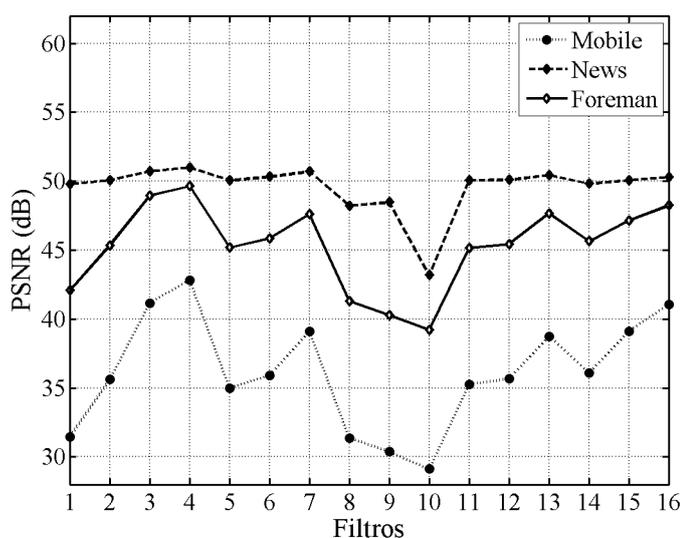


Figura 5.13 Gráfico do PSNR para dos vídeos codificados após a transcodificação.

Em relação aos resultados da métrica SSIM, os erros obtidos após a codificação possuem a mesma ordem de grandeza dos erros correspondentes aos vídeos transcodificados, como pode ser visto na Tabela 5.10.

Alguns dos vídeos obtiveram resultados melhores do que os resultados encontrados para a codificação dos vídeos originais (QCIF). O resultado do método SSIM após a codificação do vídeo original Mobile foi 0,9941, enquanto que para o vídeo original News o valor foi 0,9966 e para o vídeo original Foreman foi 0,9962.

Para o vídeo Mobile os vídeos gerados pelos filtros média 4×4 obtiveram resultados melhores que os resultados do vídeo original. Para o vídeo News os vídeos gerados pelos filtros média 3×3 , média 4×4 , mediana 3×3 , mediana 4×4 , média ponderada 3 e sigma 4×4 obtiveram resultados melhores que os resultados do vídeo original. Para o vídeo Foreman os vídeos gerados pelos filtros média 4×4 obtiveram resultados melhores que os resultados do vídeo original.

Comparando a Tabela 5.10 com a Tabela 5.7 é visto que para o vídeo News todos os resultados da métrica SSIM obtiveram resultados melhores que para os vídeos transcodificados. Para o vídeo Foreman apenas os vídeos gerados com os filtros mediana 2×2 e sigma 2×2

Tabela 5.10 Resultados dos vídeos codificados após a transcodificação utilizando o método SSIM.

Numeração	Filtros	Mobile	News	Foreman
1	Eliminação Simples	0,9719	0,9962	0,9854
2	Média 2×2	0,9849	0,9962	0,9921
3	Média 3×3	0,9940	0,9969	0,9961
4	Média 4×4	0,9951	0,9972	0,9966
5	Mediana 2×2	0,9829	0,9962	0,9918
6	Mediana 3×3	0,9832	0,9966	0,9929
7	Mediana 4×4	0,9893	0,9970	0,9950
8	Moda 2×2	0,9700	0,9949	0,9827
9	Moda 3×3	0,9618	0,9952	0,9787
10	Moda 4×4	0,9456	0,9885	0,9747
11	Média ponderada 1	0,9844	0,9962	0,9919
12	Média ponderada 2	0,9852	0,9963	0,9924
13	Média ponderada 3	0,9910	0,9966	0,9951
14	Sigma 2×2	0,9863	0,9961	0,9925
15	Sigma 3×3	0,9909	0,9963	0,9944
16	Sigma 4×4	0,9928	0,9967	0,9954

obtiveram resultados melhores na transcodificação quando comparado com a codificação. Para o vídeo Mobile os vídeos gerados utilizando os filtros moda, média ponderada 1 e 2 e sigma 2×2 obtiveram resultados melhores na transcodificação quando comparado com a codificação.

A correlação entre as medidas da SSIM dos vídeos transcodificados e dos vídeos transcodificados espacialmente após a codificação mostra que para o vídeo Foreman a correlação é fraca, com 0,3792. Para os vídeos Mobile e News a correlação entre as medidas é bem fraca, com 0,105 e -0,0213, respectivamente.

Na Figura 5.14 é visto que quanto maior o movimento maior é o erro gerado na codificação e que os vídeos usando os filtros da moda e da eliminação simples obtiveram os piores resultados.

5.4 Testes Subjetivos dos Vídeos Transcodificados

Para a realização dos testes subjetivos foram escolhidos os oito melhores resultados, avaliando os resultados encontrados pelos métodos PSNR e SSIM. Os vídeos selecionados foram os que usaram os seguintes filtros: sigma 2, mediana 2, média 3, média ponderada 3, sigma 3, média ponderada 1, média ponderada 2 e a mediana 3.

Os testes subjetivos foram realizados usando o método PC, o dispositivo N95 e todos os vídeos codificados usando o codificador H.264 com taxa de *bit* de 243 *kbit/s*. Os oito vídeos foram divididos em dois grupos de quatro vídeos para a realização dos testes. Essa divisão ocorreu para seguir a norma ITU-R BT.500-11 (ITU-R, 2002), que diz que os teste não devem passar de 30 minutos de vídeos. Para cada grupo foram utilizados os mesmos três vídeos utilizados na Seção 5.3.

Vídeos	Nome
1	Eliminação Simples
2	Média 2×2
3	Média 3×3
4	Média 4×4
5	Mediana 2×2
6	Mediana 3×3
7	Mediana 4×4
8	Moda 2×2
9	Moda 3×3
10	Moda 4×4
11	Média ponderada 1
12	Média ponderada 2
13	Média ponderada 3
14	Sigma 2×2
15	Sigma 3×3
16	Sigma 4×4

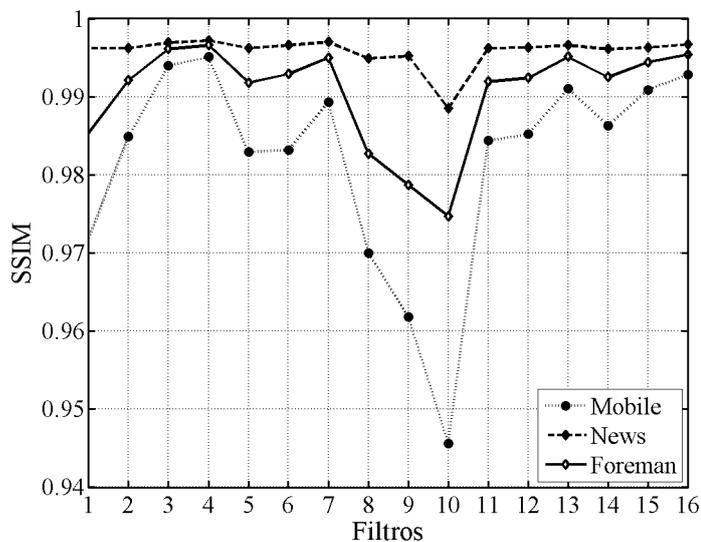


Figura 5.14 Gráfico do SSIM para os vídeos codificados após a transcodificação.

5.4.1 Avaliação do Primeiro Grupo de Vídeos

O primeiro grupo de quatro vídeos é composto pelos vídeos gerados usando os seguintes filtros: sigma 2, média ponderada 3, sigma 3 e mediana 3. Os resultados do MOS, do desvio padrão e do coeficiente de variação para o primeiro grupo de teste subjetivos são apresentados na Tabela 5.11 e na Figura 5.15.

Tabela 5.11 Valores do MOS, desvio padrão e CV dos dados coletados do primeiro grupo de testes subjetivos para os vídeos transcodificados.

Vídeos	Nome	MOS	Desvio Padrão	CV %
1	Foreman sigma2	7,7899	0,1026	1,3168
2	Foreman mediana3	7,5000	0,1201	1,6014
3	Foreman ponderada3	7,5000	0,1207	1,6092
4	Foreman sigma3	7,3083	0,1333	1,8234
5	Mobile sigma2	7,3167	0,1147	1,5681
6	Mobile mediana3	7,5042	0,1291	1,7964
7	Mobile ponderada3	7,5630	0,1161	1,5351
8	Mobile sigma3	7,1849	0,1257	1,6746
9	News sigma2	7,7500	0,1099	1,4182
10	News mediana3	7,2231	0,1325	1,8346
11	News ponderada3	7,3140	0,1379	1,8852
12	News sigma3	7,2917	0,1382	1,8955

Os testes subjetivos foram realizados com 20 participantes entre homens e mulheres com faixa de idade entre 15 e 52 anos, todos não especialistas da área de vídeo. Os testes duraram

em média 32 minutos, contando o tempo necessário para a instrução do teste e o tempo para colocar os vídeos na tela. Os participantes forneceram valores discretos de 0 a 10, para cada vídeo assistido, correspondendo ao seu julgamento da qualidade do vídeo. Cada participante assistiu seis vezes cada vídeo gerando assim 120 amostras por vídeo.

Para todos os vídeos o desvio padrão obtido teve um valor considerado pequeno, variando entre 0,1026 e 0,1382. Em relação ao coeficiente de variação, é possível verificar que todos os resultados obtidos são ditos de dispersão baixa. O coeficiente de variação variou entre 1,3168 e 1,8955, tendo uma pequena variação entre o máximo e o mínimo.

Os vídeos Foreman e News obtiveram resultados com o mesmo comportamento, com o vídeo transcodificado usando o filtro sigma 2×2 obtendo o melhor MOS, como pode ser visto na Figura 5.15, 7,7899 para o vídeo Foreman e 7,75 para o vídeo News. Para o vídeo Mobile os vídeos transcodificados espacialmente utilizando os filtros média ponderada 3 e sigma 3×3 obtiveram os melhores MOS, na qual para o da média ponderada 3 obteve 7,563 e o sigma 3 7,5042.

Vídeos	Nome
1	Foreman sigma2
2	Foreman mediana3
3	Foreman ponderada3
4	Foreman sigma3
5	Mobile sigma2
6	Mobile mediana3
7	Mobile ponderada3
8	Mobile sigma3
9	News sigma2
10	News mediana3
11	News ponderada3
12	News sigma3

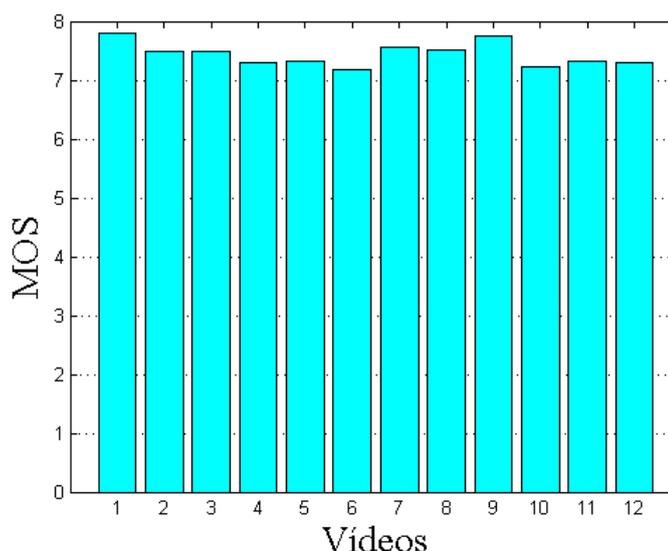


Figura 5.15 Valores do MOS para o primeiro grupo de vídeos transcodificados.

Com os dados do MOS foi possível encontrar a correlação com as medidas da PSNR e do SSIM. A correlação entre a PSNR e o MOS encontrada foi bem fraca para os vídeos Foreman e Mobile, com 0,1020 e 0,0019 respectivamente. Para o vídeo News a correlação encontrada foi muito forte, obtendo 0,9972. A correlação entre a SSIM e o MOS encontrada foi muito forte para os vídeos Foreman e News, com 0,9476 e 0,9356 respectivamente. Para o vídeo Mobile a correlação encontrada foi bem fraca, obtendo -0,1314.

Os resultados mostrados da correlação comprovam que o método SSIM tem uma melhor correlação com os testes subjetivos quando comparado com o método PSNR. Mas é possível observar que dependendo do vídeo o método SSIM não tem uma boa relação com os testes subjetivos.

5.4.2 Avaliação do Segundo Grupo de Vídeos

O segundo grupo de quatro vídeos é composto pelos vídeos gerados usando os seguintes filtros: mediana 2×2 , média 3×3 , média ponderada 2 e média ponderada 1. Os resultados do MOS, do desvio padrão e do coeficiente de variação para o segundo grupo de teste subjetivos são apresentados na Tabela 5.12 e na Figura 5.16.

Tabela 5.12 Valores do MOS, desvio padrão e CV dos dados coletados do segundo grupo de testes subjetivos para os vídeos transcodificados.

Vídeos	Nome	MOS	Desvio Padrão	CV %
1	Foreman mediana2	7,5630	0,1003	1,3263
2	Foreman média3	7,3644	0,1101	1,4949
3	Foreman ponderada2	7,2857	0,1014	1,3913
4	Foreman ponderada1	7,2250	0,1014	1,4039
5	Mobile mediana2	7,3025	0,0944	1,2926
6	Mobile média3	7,4417	0,1090	1,4644
7	Mobile ponderada2	7,1513	0,0960	1,3417
8	Mobile ponderada1	7,2083	0,1003	1,3921
9	News mediana2	7,4538	0,0945	1,2683
10	News média3	7,0167	0,1258	1,7928
11	News ponderada2	7,3250	0,1053	1,4375
12	News ponderada1	7,3109	0,0930	1,2723

Os testes subjetivos foram realizados com 20 participantes entre homens e mulheres com faixa de idade entre 15 e 55 anos, todos não especialistas da área de vídeo. Os testes duraram em média 32 minutos, contando o tempo necessário para a instrução do teste e o tempo para colocar os vídeos na tela. Os participantes forneceram valores discretos de 0 a 10, para cada vídeo assistido, correspondendo ao seu julgamento da qualidade do vídeo. Cada participante assistiu seis vezes cada vídeo gerando assim 120 amostras por vídeo.

Para todos os vídeos o desvio padrão obtido teve um valor considerado pequeno, variando entre 0,0930 e 0,1258. Em relação ao coeficiente de variação, é possível verificar que todos os resultados obtidos são ditos de dispersão baixa. O coeficiente de variação variou entre 1,2683 e 1,7928, tendo uma pequena variação entre o máximo e o mínimo.

Não ocorreu uma grande variação entre os 12 vídeos avaliados para o grupo. Para os vídeos Foreman e News o melhor MOS foi obtido utilizando o filtro sigma 2×2 como apresentado na Figura 5.16, 7,5630 para o vídeo Foreman e 7,4538 para o vídeo News. Para o vídeo Mobile o vídeo transcodificado espacialmente utilizando o filtro média 3×3 obteve o melhor MOS, 7,4417.

Com os dados do MOS foi possível encontrar a correlação com as medidas da PSNR e do SSIM. A correlação entre a PSNR e o MOS encontrada foi fraca para o vídeo Foreman, com -0,3062. Para os vídeos Mobile e News a correlação encontrada foi média, obtendo 0,6296 e

Vídeos	Nome
1	Foreman mediana2
2	Foreman media3
3	Foreman ponderada2
4	Foreman ponderada1
5	Mobile mediana2
6	Mobile media3
7	Mobile ponderada2
8	Mobile ponderada1
9	News mediana2
10	News media3
11	News ponderada2
12	News ponderada1

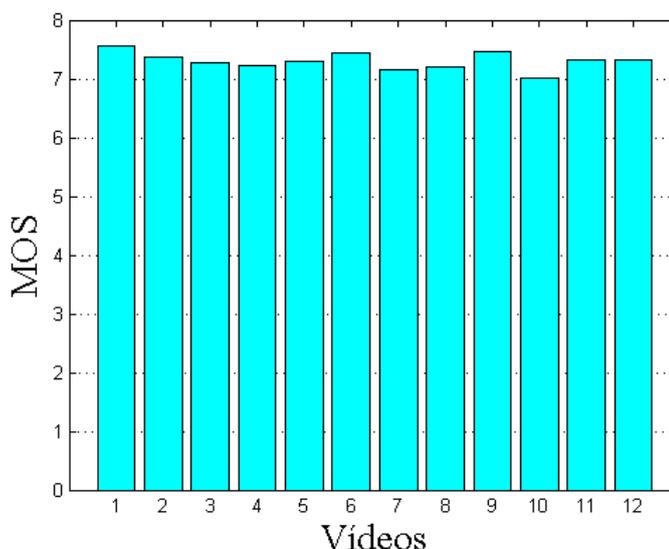


Figura 5.16 Valores do MOS para o segundo grupo de vídeos transcodificados.

0,47, respectivamente. A correlação entre SSIM e MOS encontrada foi média para os vídeos Foreman e Mobile, com 0,6190 e - 0,6880, respectivamente. Para o vídeo News a correlação encontrada foi muito forte, obtendo 0,9487.

5.4.3 Avaliação Geral dos Dados Subjetivos

Observando os resultados obtidos nos testes subjetivos dos dois grupos de vídeos, é possível analisar os filtros para cada um dos três vídeos. Assim essa Subseção é formada pela avaliação do MOS dos vídeos transcodificados espacialmente do primeiro e segundo grupo, separados entre os vídeos Foreman, Mobile e News.

Os participantes do teste subjetivo não conseguiram verificar uma real diferença entre os vídeos, como pode ser verificado ao analisar os valores do MOS em relação ao desvio padrão para todos os vídeos.

Para o vídeo Foreman os valores do MOS são apresentados na Figura 5.17. O melhor resultado para esse vídeo foi o vídeo transcodificado utilizando o filtro sigma 2×2 , logo após se destaca os vídeos transcodificados com os filtros mediana 2×2 , média ponderada 3 e mediana 3×3 .

Para o vídeo Mobile os valores do MOS são apresentados na Figura 5.18. Os melhores resultados para esse vídeo foram os vídeos transcodificados utilizando o filtro média ponderada 3 e mediana 3×3 .

Para o vídeo News os valores do MOS são apresentados na Figura 5.19. Os melhores resultados para esse vídeo foram os vídeos transcodificados utilizando o filtro sigma 2×2 e mediana 2×2 .

A correlação entre os resultados do MOS e da PSNR para cada vídeo foi calculada, resultado em uma correlação fraca para os vídeos Foreman e Mobile, 0,3721 e 0,3209, respec-

Vídeos	Nome
1	Sigma 2
2	Mediana 2
3	Média 3
4	Média Ponderada 3
5	Sigma 3
6	Média Ponderada 2
7	Média Ponderada 1
8	Mediana 3

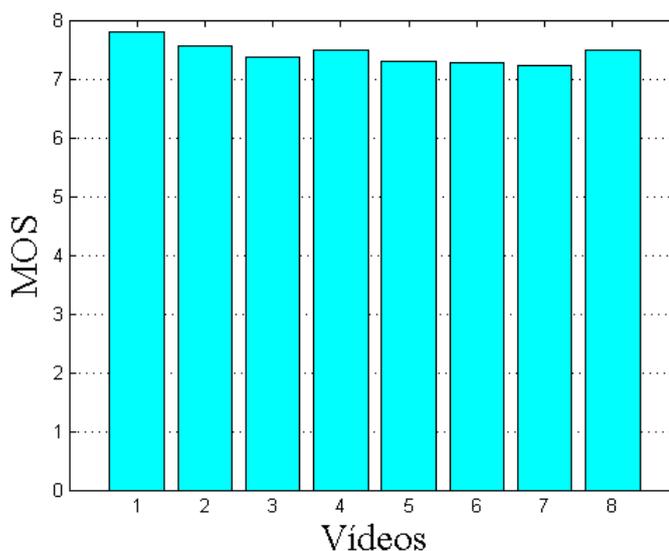


Figura 5.17 Valores do MOS para o vídeo Foreman.

Vídeos	Nome
1	Sigma 2
2	Mediana 2
3	Média 3
4	Média Ponderada 3
5	Sigma 3
6	Média Ponderada 2
7	Média Ponderada 1
8	Mediana 3

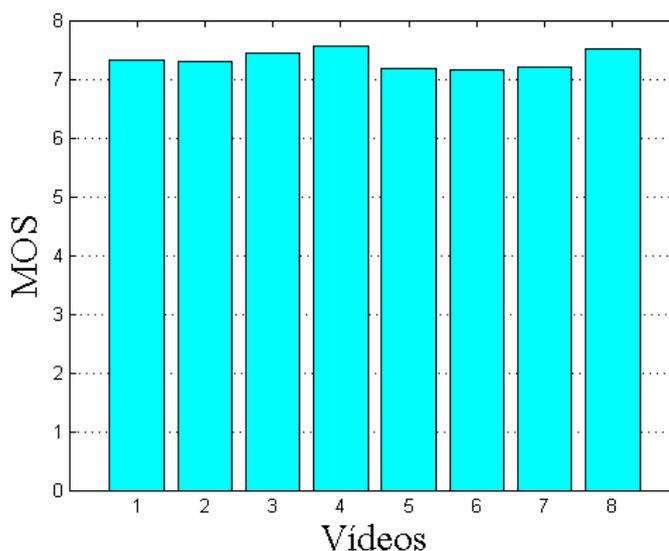


Figura 5.18 Valores do MOS para o vídeo Mobile.

tivamente, e uma correlação forte para o vídeo News, 0,7745. Já a correlação entre o MOS e o SSIM obteve valores melhores, como esperado. Para o vídeo Foreman a correlação entre o SSIM e o MOS é média, 0,5837, para o vídeo Mobile a correlação é fraca, - 0,372 e para o vídeo News a correlação é forte, 0,8486.

Com esses resultados do MOS verifica-se que para os vídeos utilizados, os filtros que obtiveram os melhores resultados foram o sigma 2×2 , mediana 3×3 , média ponderada 3 e mediana 2×2 .

Vídeos	Nome
1	Sigma 2
2	Mediana 2
3	Média 3
4	Média Ponderada 3
5	Sigma 3
6	Média Ponderada 2
7	Média Ponderada 1
8	Mediana 3

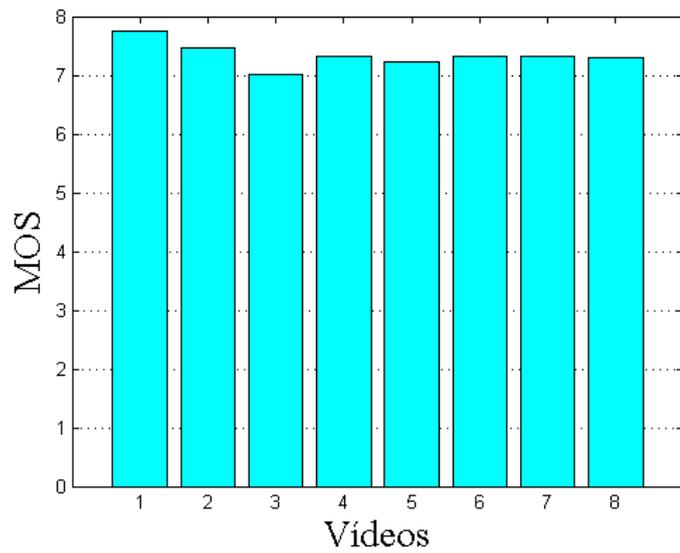


Figura 5.19 Valores do MOS para o vídeo News.

CAPÍTULO 6

Conclusões

Este trabalho descreve um transcodificador espacial de vídeo para a TV móvel e avaliações de determinados parâmetros utilizados. Os vídeos transcodificados, bem como os parâmetros da TV móvel foram avaliados de acordo com medidas objetivas e subjetivas. A avaliação objetiva foi realizada utilizando os métodos PSNR e SSIM, enquanto a avaliação subjetiva foi realizada utilizando o método PC.

Na codificação dos vídeos tem-se uma diminuição no erro provocado pelo codificador à medida que se aumenta a janela utilizada no filtro, uma vez que resulta em vídeos com muita redundância espacial, como demonstrada na avaliação objetiva.

Utilizando métodos subjetivos foram avaliados dois dispositivos, que podem ser usados para a TV móvel com mesma taxa de *bit*. Os dispositivos testados foram o celular NOKIA N95 e o PDA NOKIA N800. Foram obtidos valores de MOS mais altos nos testes do dispositivo N95 do que os obtidos nos testes do dispositivo N800, em razão do tamanho da tela, que é menor no celular.

Para cada um dos vídeos, foram testadas três diferentes taxas de *bits*: 111 *kbit/s*; 243 *kbit/s* e 318 *kbit/s*. Os testes mostraram que as taxas de 243 *kbit/s* e 318 *kbit/s* obtiveram resultados de MOS melhores do que a taxa de 111 *kbit/s*. Entretanto, não se tem uma percepção significativamente diferente entre elas, o que torna a taxa de 243 *kbit/s* mais adequada para a transmissão da TV móvel pela rede UMTS, visto que permite que mais usuários tenham acesso à rede. Essa conclusão também é válida para a TV Digital, pois permite além da transmissão do sinal de TV, maior quantidade de dados transmitidos.

A transcodificação espacial de vídeos foi apresentada neste trabalho utilizando os filtros da média, mediana, moda, sigma e média ponderada. Em relação ao tempo de processamento para a realização da transcodificação, o aumento da janela gera um aumento no tempo da transcodificação. Os filtros da média e média ponderada obtiveram valores mais baixos, enquanto que o da moda e sigma valores mais altos.

Foram apresentados ainda os erros provocados pela codificação H.264. Para alguns dos vídeos transcodificados espacialmente, esses erros foram menores que os provocados para os

vídeos originais. Exemplo disso é o vídeo transcodificado utilizando o filtro da média 4×4 , que obteve, para todos os vídeos, um erro menor que o original. Situação análoga para os vídeos News, transcodificados utilizando os filtros média 3×3 , mediana 3×3 , mediana 4×4 , média ponderada 3 e o sigma 4×4 .

Ao se analisar o erro provocado pela transcodificação nos três vídeos, os resultados da medida objetiva PSNR apresentam o filtro sigma 2×2 como o de melhor qualidade, enquanto que os resultados da medida objetiva SSIM apresentam os filtros mediana 2×2 e sigma 2×2 como os de melhores desempenhos.

Nos testes subjetivos, os vídeos transcodificados espacialmente utilizando os filtros mediana 2×2 e sigma 2×2 obtiveram melhores resultados para os vídeos Foreman e News. Para o vídeo Mobile, os vídeos utilizando os filtros da média ponderada 3 e média 3×3 obtiveram as melhores qualidades.

É apresentada ainda a correlação entre as medidas objetivas e subjetivas, com a correlação entre a medida SSIM e MOS obtendo valores superiores aos da correlação entre as medidas PSNR e MOS.

Comparando as medidas subjetivas e objetivas, os vídeos transcodificados utilizando os filtros mediana 2×2 e sigma 2×2 são os que apresentam melhor desempenho. O filtro da mediana 2×2 possui pequena vantagem em relação ao tempo necessário para o processamento da transcodificação.

Sobre a aceitação da TV móvel, foi possível notar (no questionário, Apêndice C) que existe uma boa expectativa sobre seu uso, visto que 90% dos entrevistados gostariam de ter o serviço à sua disposição. Um outro dado importante é a frequência diária de utilização. Para mais de 50% dos participantes esse serviço seria utilizado mais de duas vezes por dia, o que demonstra forte interesse pelo serviço.

Sugestões para Trabalhos Futuros

Como possibilidades de trabalhos futuros, destaca-se a transcodificação de resoluções espaciais superiores, como HDTV e SDTV, para resoluções espaciais que possam ser utilizadas na TV móvel, como CIF e QCIF.

É viável ainda a inserção de ruídos de transmissão nos vídeos originais e transcodificados, visando avaliar a degradação na qualidade dos vídeos transcodificados espacialmente em relação aos vídeos originais.

O grupo VQEG disponibiliza vídeos com ruídos, para simular vídeos que sofreram degradação antes da transmissão. Assim, o modelo de transcodificação apresentado neste trabalho poderá ser uma forma de diminuir a degradação desses vídeos antes de sua transmissão.

Alguns parâmetros ainda precisam ser definidos e uma outra possibilidade seria a realização de testes subjetivos para identificar a resolução espacial de vídeos (CIF, QVGA, QCIF), bem como a resolução temporal de preferência da população brasileira.

APÊNDICE A

Medidas Subjetivas

A recomendação BT.500 (ITU-R, 2002) especifica várias características que devem ser considerados nos testes, como mostrado abaixo.

- Número de indivíduos: pelo menos 30, e de preferência mais. Eles deveriam ter visão normal, e deve preferencialmente ser não-especialista. Cada indivíduo é caracterizado pela idade, sexo e profissão. Todos devem passar por um teste de acuidade visual e percepção de cores. Para cada indivíduo está associado um número de identificação, ao qual está associadas informações sobre o indivíduo e sessão de testes (fabricante e modelo do monitor utilizado, distância de visualização etc).
- Teste cenas: estas devem ser críticas para avaliação das imagens sob teste.
- Teste sessão: a parte experimental não deveria ter mais de 30 minutos devido à possibilidade de cansaço ao se avaliar as imagens. No total, com a instrução parte do ensaio não deve ser mais de 1 hora.
- Condições: especificações foram estabelecidas para o ambiente, iluminação e da distância de visualização. Dois tipos de ambientes são permitidos, laboratório e doméstico. Mesmo em ambiente doméstico, condições de contorno estreitas são aplicadas, de forma a assegurar que os resultados possam ser reproduzidos posteriormente.
- Apresentação estímulo: apresentação aleatória das imagens ou das sequências é recomendado.

A.1 Medidas Subjetivas para Aplicações em Televisão

Dentre os principais métodos de avaliação para TV tem-se:

- Método DSIS - *Double Stimulus Impairment Scale*;

- Método DSCQS - *Double Stimulus Continuous Quality Scale*;
- Método SS - *Single Stimulus*;
- Métodos de Comparação de Estímulos (*Stimulus Comparison Method*);
- Método SSCQE - *Single Stimulus Continuous Quality Evaluation*

A.1.1 Método DSIS

É comumente usado para se ter a percepção de defeitos ocorridos em percursos de transmissão e para avaliação de novos sistemas. O avaliador é primeiramente apresentado a um conjunto de imagens de referência, em que não ocorrem artefatos ou outros defeitos, e em seguida, ao mesmo conjunto após seu processamento pelo sistema sobre teste. A ele é pedido julgar a segunda, tendo como base a primeira. O esquema do teste prevê tempos definidos para cada apresentação, assim como tempo de recuperação e descanso durante a transição (imagem cinza), e um tempo final para o julgamento.

A primeira variante dá uma única apresentação da referência e uma única da imagem a ser avaliada; a segunda variante repete as duas apresentações antes do julgamento, conforme a Figura A.1.

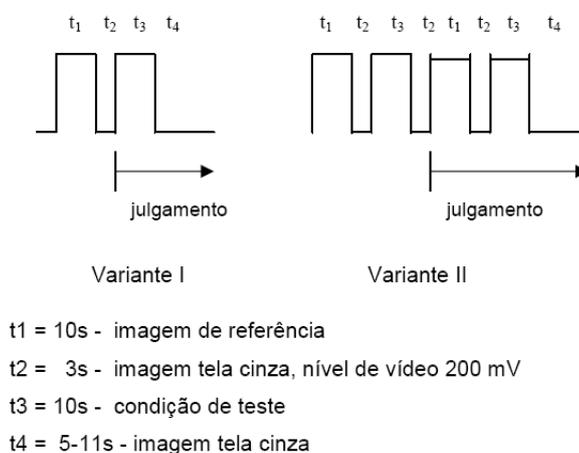


Figura A.1 Tempos de observação e julgamento, método DSIS.

O julgamento deve ser dado em termos de notas a respeito da degradação da imagem, seguindo a escala da Tabela A.1.

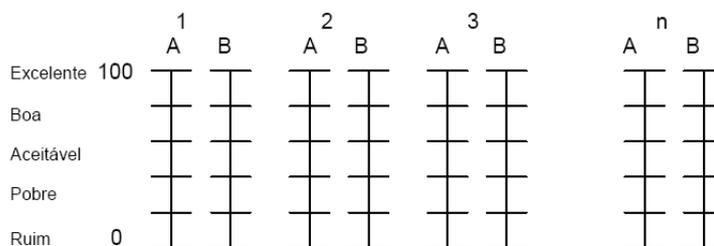
Os observadores são orientados a julgar sempre sobre a impressão geral que tiveram da última sequência. Existem formulários para que cada um responda sobre as diversas sequências a ele apresentadas.

Tabela A.1 conceito para avaliação DSIS.

conceito	degradação
5	imperceptível
4	perceptível mas não irritante
3	levemente irritante
2	irritante
1	muito irritante

A.1.2 Método DSCQS

Este método também é especialmente útil na avaliação dos efeitos dos meios de transmissão sobre a qualidade da imagem. Nele, o avaliador assiste imagens aos pares, da mesma fonte, sendo uma proveniente diretamente desta fonte e outra sendo processada pelo sistema sobre averiguação. A ordem temporal das imagens durante as diversas sequências é alterada de forma pseudo-aleatória, e não é dado o conhecimento prévio aos avaliadores de qual imagem é a referência. Ambas as imagens de cada sequência devem receber julgamento. A pessoa ou grupo avaliador vê cada uma das duas imagens em torno de dez segundos por duas ou três vezes, memoriza a percepção e então dá a nota correspondente a cada uma delas, entre os limites de 0 a 100. Lembrando que, para cada par de imagens de teste, uma não apresenta imperfeições, enquanto que a outra pode ou não contê-las. O observador deve inserir uma marca na escala vertical contínua da Figura A.2 para cada uma das imagens apresentadas aos pares.

**Figura A.2** Classificação da qualidade da imagem pelo método DSCQS.

Não é conveniente se associar os números da pontuação dada neste método com os resultados e adjetivos do método anterior. Os resultados obtidos devem ser tratados apenas como a diferença de classificação numérica entre a referência e o sistema sobre teste (ITU-R, 2002). Desta forma, um número pequeno na escala DSCQS indica que a diferença de qualidade da imagem entre a fonte e o equipamento ensaiado é pequena, ou seja, o sistema introduz pouca degradação. Porém, não indica a qualidade final do sistema.

A.1.3 Método SS

Neste caso, uma única sequência de imagens é apresentada, e de forma dirigida para se avaliar um determinado defeito ou sintoma. É interessante para se obter uma gama de valores para um mesmo tipo de degradação. O julgamento pode ser dado de duas formas:

- a) Por notas correspondentes aos adjetivos (5 = excelente, ou imperceptível; 4 = boa ou perceptível, mas não irritante; 3 = aceitável, ou levemente irritante; 2 = pobre, ou irritante; 1 = ruim, ou muito irritante). Permite-se a utilização de meio grau dentro da escala, especialmente quando se quer examinar detalhes, como a legibilidade de textos na imagem, por exemplo;
- b) Método do desempenho: a percepção do avaliador é obtida com a resposta a questões a ele formuladas. Por exemplo, a identificação de um certo objeto na cena, a leitura de um texto, etc. A velocidade da resposta e o acerto são dados que formam distribuições, e a análise de resultados é baseada na tendência central e na dispersão destas distribuições.

A.1.4 Métodos de Comparação de Estímulos

Duas seqüências de imagens de mesmo conteúdo – uma referência e outra sobre teste – são mostradas simultaneamente, e o observador deve dar uma nota correspondente à relação entre as imagens. Pode-se usar dois monitores igualmente ajustados ou ainda dividir uma tela de maiores proporções. Nos testes para televisão, há três metodologias de avaliação:

- a) Julgamento com adjetivos: a relação entre as imagens é feita com uso de palavras que expressam a existência e direção de diferenças perceptíveis, de acordo com a Tabela A.2;

Tabela A.2 Conceito para avaliação no método “Comparação de estímulos”.

nota	Imagem teste comparada à referência
-3	muito pior
-2	pior
-1	levemente pior
0	igual
1	levemente melhor
2	melhor
3	muito melhor

- b) Julgamento sem categoria: a relação da qualidade observada é informada por um número, entre limites abertos ou pré-definidos, ou ainda pelo desenho de um ponto ou risco em uma escala vertical com dois limites rotulados nas pontas desta escala (exemplo: igual, diferente);

- c) Método do desempenho: cada imagem do par tem graus diferentes de um determinado sintoma, ou uma delas não o contém. O avaliador decide qual das imagens contém o problema, ou em qual delas está a de maior grau. O acerto e velocidade na resposta são usados como índices da relação entre as duas imagens.

A introdução da compressão na TV digital produz defeitos que afetam a qualidade da imagem e que são muito dependentes das cenas e suas variações; a qualidade flutua muito durante um programa longo. As metodologias até aqui citadas não são suficientes para se medir este tipo de situação. Além disto, os métodos de estímulo duplo são possíveis em laboratório, e não replicam as condições de percepção domésticas, tipicamente de estímulo único. Para a avaliação da qualidade do sinal de forma contínua, sem a imagem referencial, o método a seguir é mais apropriado.

A.1.5 Método SSCQE

Ao grupo de pessoas é apresentada uma série de segmentos de programa, cada um com diferentes parâmetros de qualidade em avaliação, tendo cada segmento duração mínima de cinco minutos. Os avaliadores têm em mãos um mecanismo sensor com um cursor linear de 10 cm, e enquanto assistem, transmitem suas impressões por meio do deslocamento do cursor em dimensão proporcional à sensação percebida. Esses valores são gravados na cadência de duas amostras por segundo, e os dados obtidos permitem o levantamento de histogramas. É uma escala contínua, como a utilizada no método DSCQS, porém, se naquele o grau era único no final de uma apresentação curta, nesse as notas são dadas de forma contínua durante o tempo da demonstração.

Esse método é mais adequado, pois é natural que em dez segundos, como nos ensaios anteriores, não haja uma representação significativa do que acontece em um caso real, mais longo, já que os artefatos são muito dependentes do conteúdo temporal e espacial. Esse método está mais adequado para medidas de qualidade de vídeo em sequências longas, e reproduz situações reais com a ausência da referência.

Porém, quando se quer medir a fidelidade, então a referência deve ser introduzida. Assim, foi proposto o próximo método.

A.1.6 Método SDSCE

As sequências de referência e de teste são apresentadas simultaneamente, em um mesmo monitor dividido, ou em dois monitores ajustados de forma igual. O grupo deve verificar as diferenças entre as duas e julgar a fidelidade do vídeo movendo a alavanca do mecanismo sensor, semelhante ao caso antes exposto. Para representar fidelidade perfeita (codificada como 100) a alavanca deve estar no máximo comprimento, e fidelidade nula (escala 0) deve estar no fundo.

É importante observar que estudos identificaram que a memória humana altera significativamente o resultado de testes como os sugeridos anteriormente. As imagens mais lembradas

de uma sequência serão as dos últimos dez a quinze segundos. O ITU considera essa questão, usando trechos curtos quando o julgamento é posterior à apresentação, enquanto que trechos longos são medidos durante sua realização (REHME, 2007).

A.2 Medidas Subjetivas para Aplicações Multimídia

Todas as formas de avaliação apresentadas anteriormente são da recomendação denominada ITU-R BT.500 - 11, que se aplica à televisão em definição convencional. Há recomendações específicas para TV de alta definição, videoconferência e outras aplicações multimídia. Outros grupos também apresentaram algumas metodologias, mas essencialmente possuem princípios similares e, dada a abrangência e aceitação das recomendações ITU, considera-se suficiente o entendimento do referido nome.

Essas medidas são feitas baseadas em opiniões dos participantes, com diferentes idades, formações profissionais e que não sejam da área do projeto. Normalmente são realizadas durante o desenvolvimento ou padronização de um novo algoritmo ou de um novo sistema. O objetivo desses testes é a avaliar o desempenho das novas ferramentas, a fim de otimizar os algoritmos ou os sistemas que estão em estudo.

Diversos métodos experimentais têm sido validados para diferentes fins. Na norma ITU-T P.910 (ITU-T, 1999), três métodos são recomendados para aplicações multimídia, sendo especificados os vídeos por classes e os vídeos usados nos testes, como vídeos com codificação de vídeo PCM de 8 *bit/pixel* em 4:2:2 e no formato Y, C_R e C_B .

A escolha final de um dos métodos para uma determinada aplicação depende de vários fatores, tais como o contexto, o objetivo e, quando no processo de desenvolvimento do ensaio, deve ser realizada. Os três métodos são: ACR (*Absolute Category Rating*), DCR (*Degradation Category Rating*), PC (*Pair Comparison*).

A.2.1 Método ACR

O método ACR é uma categoria em que as sequências do teste são apresentadas uma de cada vez e são classificadas com independência em relação a uma escala específica, esse método também é chamado Método Único de Estímulos.

O método especifica que, após cada apresentação, os entrevistados são convidados a avaliar a qualidade da sequência indicada. O período de votação deve ser inferior ou igual a 10 s, dependendo do mecanismo de votação utilizado. O tempo de apresentação pode ser reduzido ou aumentado, de acordo com o conteúdo do teste.

São utilizados os cinco níveis seguintes para a escala de classificação de qualidade global:

Se é necessário uma discriminação maior, uma escala de nove níveis pode ser utilizada. O aumento da escala é útil para a obtenção de mais informações sobre diferentes fatores

- 5 Excelente
- 4 Bom
- 3 Legível
- 2 Ruim
- 1 Péssimo

perceptíveis da qualidade quando os estímulos são quase iguais, apesar de os sistemas serem claramente percebidos como diferentes.

Para o método ACR, o número necessário de repetições é obtido mediante a realização dos testes nas mesmas condições em tempos diferentes.

A.2.2 Método DCR

O método DCR implica que as sequências do teste são apresentadas em pares: o primeiro estímulo apresentado em cada par é sempre a fonte de referência, enquanto o segundo estímulo é o sob teste, este método também é chamado de duplo estímulo.

O tempo padrão para a apresentação do estímulo é ilustrado na Figura A.3. O período de votação deve ser inferior ou igual a 10 s, dependendo do mecanismo de votação utilizado. O tempo de apresentação pode ser reduzido ou aumentado, de acordo com o conteúdo do teste.

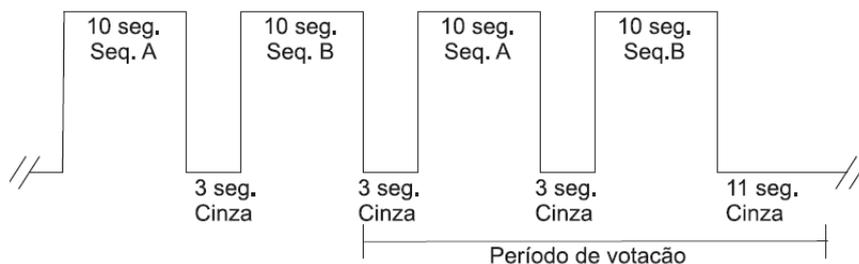


Figura A.3 Apresentação dos estímulos no método DCR.

Neste caso, os indivíduos são convidados a avaliar o comprometimento do segundo estímulo em relação ao de referência. Devem ser utilizados os cinco de níveis seguintes para a escala de classificação de qualidade global:

- 5 Imperceptível
- 4 Perceptível mas não irritante
- 3 Levemente irritante
- 2 Irritante
- 1 Muito irritante

Para o método DCR, o número necessário de repetições é obtido mediante a repetição dos testes nas mesmas condições em tempos diferentes.

APÊNDICE B

O Codificador H.264

O padrão ITU-T H.264 / MPEG-4 (Parte 10) *Advanced Video Coding* (ITU-T; 1, 2003a) (frequentemente mencionado como H.264/AVC) é uma das mais recentes propostas para codificação de vídeo digital. Foi desenvolvido pelo consórcio JVT (*Joint Video Team*) formado por especialistas oriundos do *Video Coding Experts Group* (VCEG) do ITU-T e de especialistas do *Motion Picture Expert Group* (MPEG) do ISO/IEC. Assim como nos padrões passados, o H.264/AVC propõe uma alternativa visando o equilíbrio entre eficiência de codificação, complexidade de implementação e custo, baseando-se no estado da arte de tecnologias VLSI (CPU, DSP, ASIC, FPGA). O padrão H.264 apresenta um ganho de eficiência de codificação em relação ao MPEG-2 em pelos menos duas vezes, sem aumento significativo no custo final da tecnologia. Por esta razão, o H.264 é o padrão de compressão geralmente utilizado para transmissão para receptores móveis, sendo o codificador utilizado neste trabalho para redução de taxa de *bits* (ALENCAR, 2007).

Em julho de 2004 foi adicionada ao padrão H.264/AVC uma nova parte chamada *Fidelity Range Extensions* (FRExt) (ITU-T; 1, 2003b) que especifica melhorias na eficiência de codificação, visando aplicações para alta definição, produção e edição de vídeo digital. Nas Figuras B.1 e B.2, são apresentados os diagramas de blocos do codificador e decodificador genéricos do H.264/AVC, respectivamente.

Semelhantemente aos padrões anteriores (MPEG-1, MPEG-2 e H.263 (ITU-T, 2000)), o padrão H.264/AVC não define explicitamente um CODEC (COdificador/DECodificador). O que realmente é definido na documentação é a sintaxe de uma sequência de *bits* com informação de vídeo codificado e os métodos de decodificação desta sequência de *bits*. Na prática, no entanto, um codificador e um decodificador padrão seguem as estruturas dispostas na Figura B.1 e na Figura B.2 respectivamente. Ao mesmo tempo em que todas essas funções são necessárias, ainda há espaço para modificações consideráveis na estrutura do CODEC.

Para um CODEC H.264/AVC os elementos funcionais básicos (predição, transformação, quantização e codificação de entropia) são sensivelmente diferentes dos demais padrões

anteriores. Foram essas significativas mudanças que garantiram os ganhos em eficiência de codificação alcançados pelo padrão H.264/AVC.

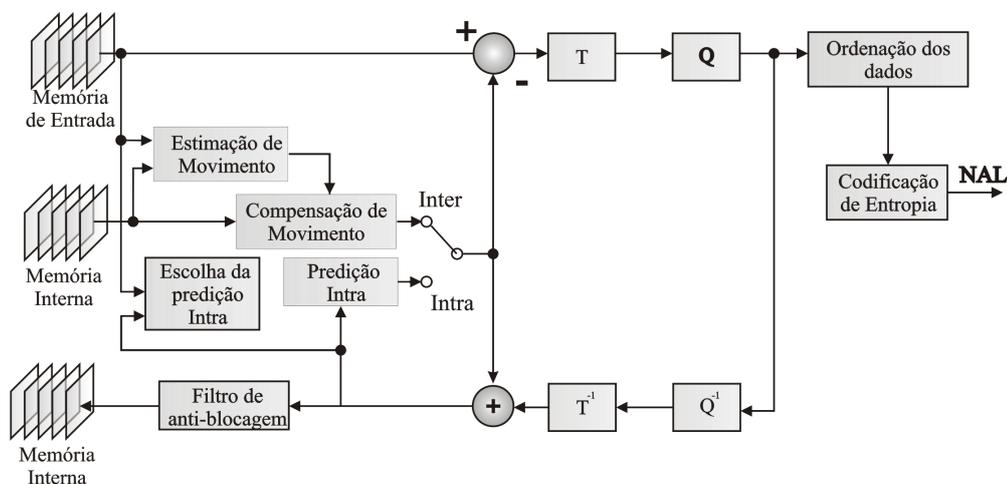


Figura B.1 Estrutura completa de um codificador H.264/AVC.

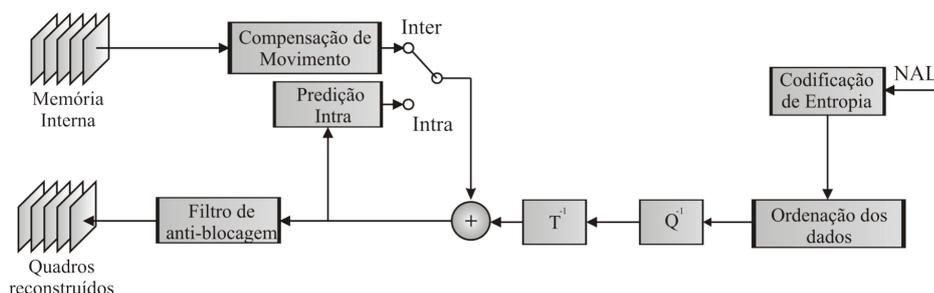


Figura B.2 Estrutura completa de um decodificador H.264/AVC.

A implementação do padrão H.264/AVC, conhecida como x264, é utilizada para desenvolvimento deste trabalho, e trata-se de uma biblioteca gratuita para codificação de seqüências de vídeo digital (MERRIT *et al.*, 2007). O x264 é usado por se tratar de uma implementação consolidada com suporte a subamostragem de crominância 4:2:0. Outra funcionalidade que foi bastante útil a este trabalho foi o relatório de saída gerado ao final de cada execução. Dados como relação sinal de pico/ruído da componente de luminância e taxa de *bits* são especialmente relevantes para a análise do impacto dos métodos estudados.

Em seguida apresenta-se, sucintamente, as principais inovações do padrão H.264/AVC, que estão implementadas no codificador x264.

B.1 O Filtro de Anti-blocagem

Um processo de filtragem é aplicado a cada macrobloco decodificado com o objetivo de reduzir o artefato de blocagem na imagem, que é caracterizado pelo aparecimento das bordas

dos macroblocos na imagem (ver Figura B.3). A bloqueio acontece principalmente devido a variações do nível DC entre macroblocos vizinhos. O filtro de anti-bloqueio é aplicado na etapa de codificação, logo após a transformada inversa no codificador e antes da reconstrução e armazenamento do macrobloco estimado para futuras previsões. Na decodificação, o filtro é aplicado na etapa anterior à reconstrução e exibição do macrobloco (RICHARDSON, 2003).



Figura B.3 Comparação entre um quadro filtrado e um quadro sem filtragem.

B.2 CABAC/CAVLC

O padrão H.264/AVC possui dois algoritmos de codificação de entropia especificados, conhecidos como codificação de palavras de comprimento variável com contexto adaptativo (*Context Adaptive Variable Length Coding* – CAVLC) e codificação aritmética de contexto adaptativo (*Context Adaptive Binary Arithmetic Coding* – CABAC). O algoritmo CABAC é um método de compressão mais poderoso, comparado ao CAVLC que é o método padrão no H.264/AVC. Ele oferece um ganho em eficiência de compressão de 10% à 15%.

B.3 Tamanho de Bloco Variável / Partições de Macroblocos

O padrão H.264/AVC suporta a divisão de um macrobloco em partições de até 4×4 *pixels*, incluindo formatos com tamanho vertical e horizontal diferentes, como por exemplo, a partição 8×4 . O tamanho de bloco é adaptativo e variável. Um bom codificador é capaz de decidir o tamanho de bloco ótimo para ser usado para cada macrobloco.

B.4 Compensação de movimento com múltiplos quadros de referência

O padrão H.264/AVC possibilita codificação mais eficiente por permitir a seleção no processo de compensação de movimento entre um número maior de quadros decodificados e armazenados no caminho inverso, conforme ilustra a Figura B.4. Os vetores de movimento são

transmitidos em conjunto com os parâmetros de referência do quadro. Esse conceito é estendido também para *slices B*.

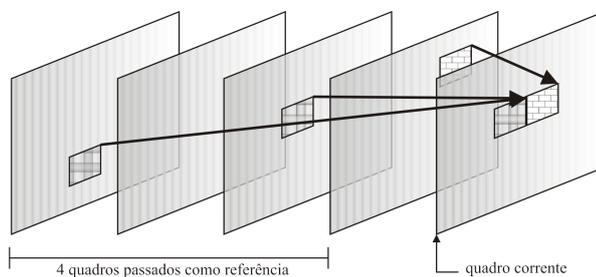


Figura B.4 Referência múltipla.

B.5 Otimização da Taxa de Distorção

A otimização da taxa de distorção permite ao codificador tomar, eficientemente, as decisões de codificação quando tiver de escolher entre diferentes opções de codificação, por exemplo, na determinação da área de busca para a compensação de movimento. A otimização da taxa de distorção não é definida nas especificações do padrão, mas está inserida no *software* de referência do H.264/AVC.

APÊNDICE C

Questionário - TV Móvel

1) Você gostaria de ter esse tipo de serviço?

Sim Não

2) Em que lugares normalmente você utilizaria esses serviços?

Bancos Restaurantes Rua Outros
 Shopping Casa Trabalho/Escola

3) Quantas vezes você utilizaria esses serviços por dia?

Nenhuma Duas vezes
 Uma vez Várias

4) Você usaria esses serviços dentro da sua casa, onde existem TVs convencionais?

Sim Não

5) Você acha inconveniente ter que fazer downloads para assistir a TV móvel?

Sim Não

6) Quanto você pagaria, em média, por esse serviço?

- 0 reais 5 reais 10 reais
 20 reais 40 reais

7) Qual serviço você escolheria?

- A TV móvel como a TV convencional, com uma programação fixa, sem custo adicional.
 A TV móvel com a programação a sua escolha com um custo adicional.

8) Você gostaria de ver a mesma programação da TV tradicional?

- Sim
 Sim, com mais opções.
 Não

9) Que tipos de programas você assistiria em uma TV Móvel?

- Notícias Novelas Séries Esportes
 Filmes Desenhos Documentários

10) Você acha importante a presença de um guia de programação?

- Sim Não

C.1 Expectativas dos Usuários sobre a TV móvel

O objetivo dessa seção é saber quais são as expectativas dos futuros usuários da TV móvel. Essas expectativas são importantes para que os ajustes necessários possam ser feitos de forma que o serviço tenha os parâmetros necessários para seus usuários. Para a avaliação foi utilizado o questionário, que contou com a participação de 50 pessoas entre 15 e 55 anos, todas não especialista e com o segundo grau completo.

Dos entrevistados, 90% gostariam de ter o serviço de TV móvel. Para 100% dos entrevistados seria necessário um guia de programação. Os locais onde esse serviço seria mais utilizado foram: bancos, com 48%, trabalho/escola, com 46%, *shopping center*, com 36%, rua, com 28%, casa, com 24% e restaurantes, com 10%.

Em relação ao número de vezes por dia em que os entrevistados optantes pelo serviço assistiriam a TV móvel, a pesquisa mostra que 52% assistiriam mais de duas vezes por dia, 22% assistiriam duas vezes por dia e 24% uma vez ao dia.

Os principais conteúdos que os entrevistados gostariam de assistir são: notícias, com 82%, documentários, com 46%, esportes e filmes, com 50%, series, com 44%, desenhos, com 32% e novelas, com 18%.

Sobre o tipo de programação 34% não estariam dispostos a assistir a mesma programação da TV tradicional, enquanto que dos 66% dos que gostariam ter a programação igual a da TV convencional, sendo 42% gostariam de ter mais opções.

Para 74% dos entrevistados é inconveniente fazer o *download* para assistir o vídeo e 60% pagariam mais de R\$ 10,00 para ter a TV móvel.

APÊNDICE D

Artigos Publicados

Artigo aceito para publicação.

1. REGIS, C. D. M.; MORAIS, D. C.; FARIAS, M. C. Q.; ALENCAR, M. S. **Objective and Subjective Assessment of Space Transcoded Videos for Mobile Receivers.** In: 2009 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, Bilbao, Spain.

Artigos publicados.

1. REGIS, C. D. M.; MORAIS, D. C.; FARIAS, M. C. Q.; ALENCAR, M. S. **Video Spatial Transcoding Assessment Based on Structural Distortion Measurement.** In: International Workshop on Telecommunications (IWT 2009), São Paulo, Brasil. 2009.
2. REGIS, C. D. M.; ROCHA, R. B.; FARIAS, M. C. Q.; ALENCAR, M. S. **Assessment of Space Transcoded Videos to Mobile Receivers Using the Average Filter.** In: 7th International Information and Telecommunication Technologies Symposium, 2008, Foz do Iguaçu, Brasil.
3. ROCHA, J. S.; REGIS, C. D. M.; FARIAS, J. E. P.; ALENCAR, M. S. **Interference Analysis of the Mobile TV Transmission over WCDMA Networks.** In: 7th International Information and Telecommunication Technologies Symposium, 2008, Foz do Iguaçu, Brasil.
4. REGIS, C. D. M.; ROCHA, J. S.; OLIVEIRA, J.; FARIAS, M. C. Q.; ALENCAR, M. S. **Transcodificação de Vídeo Digital para TV Móvel.** In: 6th International Information and Telecommunication Technologies Symposium, 2007, Brasília, Brasil.
5. REGIS, C. D. M.; OLIVEIRA, J.; ROCHA, J. S.; ANJOS G. C. B.; FARIAS, M. C. Q.; ALENCAR, M. S. **Transcodificação de vídeo digital para receptores portáteis.** In: IV SEGeT Simpósio de Excelência em Gestão e Tecnologia, 2007, Resende.

6. ALVES, M. J.; SANTOS, K. C.; REGIS, C. D. M.; ALENCAR, M. S.; PEROTONI, M. **Uma Nova Antena Fractal para Aplicação em Televisão Digital Móvel na Banda VHF.** In: XXVI Simpósio Brasileiro de Telecomunicações (SBrT'08), 2008, Rio de Janeiro.
7. ROCHA, J. S.; REGIS, C. D. M.; COSTA, P. L.; FARIAS, J. E. P.; ALENCAR, M. S. **Avaliação da Transmissão de Televisão Digital Móvel em Redes UMTS.** In: XXVI Simpósio Brasileiro de Telecomunicações (SBrT'08), 2008, Rio de Janeiro.
8. ROCHA, J. S.; REGIS, C. D. M.; COSTA, P. L.; FARIAS, J. E. P.; ALENCAR, M. S. **Análise de Desempenho de redes UMTS para Televisão Móvel.** In: MOMAG 2008, 2008, Florianópolis. p. 545-549.
9. SILVA E. F.; REGIS C. D. M.; FONTGALLAND G. **Técnica do Espalhador Opticamente Modulado nas Configurações Monoestática e Biestática.** In: MOMAG 2008, 2008, Florianópolis. p. 1018-1022.
10. COSTA M. A. D.; MEDEIROS W. G.; REGIS C. D. M.; NASCIMENTO A. R. Z.; SILVA E. F.; SOUZA M. P., ALENCAR M. S. **Interface para Auxílio à Comunicação Alternativa de Imagens e Dados.** In: V Congresso IBERDISCAP 2008, Cartagena de Indias, Colômbia.

Resumo aceito para publicação.

1. SILVA E. F.; REGIS C. D. M.; FONTGALLAND G. **Modulated Scatterer Technique: Comparison Between the Monostatic and Bistatic Configurations.** In: Progress In Electromagnetics Research Symposium, PIERS 2007, Beijing, China.

Referências Bibliográficas

ABERT. *Mais uma novela para celular será filmada no Brasil*. 2009. [Http://www.abert.org.br/D_mostra_clipping.cfm?noticia=104955](http://www.abert.org.br/D_mostra_clipping.cfm?noticia=104955). Visitado em 14 de fevereiro de 2009.

ABNT. *ABNT NBR 15602-1, Televisão digital terrestre – Codificação de vídeo, áudio e multiplexação – Parte 1: Codificação de vídeo*. Novembro 2007.

ABNT. *ABNT NBR 15604, Televisão digital terrestre – Receptores*. Novembro 2007.

ACHARYA, T.; RAY, A. K. *Image Processing - Principles and Applications*. Hoboken, New Jersey, USA: John Wiley & Sons, Inc., 2005. ISBN 13978-0-471-71998-4.

ADNEWS. *Brasileiro quer TV paga no celular*. 2008. [Http://sbtvd.cpqd.com.br/?obj=noticia&mtd=detalhe&q=12804](http://sbtvd.cpqd.com.br/?obj=noticia&mtd=detalhe&q=12804), Acessado em 29/10/2008.

AHMAD, A. M.; IBRAHIM, I. K. *Multimedia Transcoding in Mobile and Wireless Networks*. New York: Information Science Reference, 2008.

AHMAD, I. *et al.* Video transcoding: an overview of various techniques and research issues. *IEEE Transactions on Multimedia*, v. 7, n. 5, p. 793–804, Oct. 2005. ISSN 1520-9210.

AHMAD, M.; SUNDARARAJAN, D. A fast algorithm for two dimensional median filtering. *IEEE Transactions on Circuits and Systems*, v. 34, n. 11, p. 1364–1374, Nov 1987. ISSN 0098-4094.

ALENCAR, M. S. A história secreta da televisão digital no brasil(1). Artigo para o jornal eletrônico na Internet, *Jornal do Comercio On Line*, Recife, Brasil, 2005.

ALENCAR, M. S. *Televisão Digital*. São Paulo: Érica, 2007.

ARAÚJO, A. de A. *Filtros Espaciais: Estudo Comparativo e Aplicações em Classificação e Segmentação de Imagens*. Tese (Tese de Doutorado) — Universidade Federal da Paraíba - UFPB - Campus II, Campina Grande - PB, Brasil, Julho 1987.

ARTHUR, R. *Avaliação Objetiva de Codecs de Vídeo*. Dissertação (Dissertação de Mestrado) — Universidade Estadual de Campinas - UNICAMP, Campinas, Brasil, Abril 2002.

BERNADINO, F. M. *Quantização Vetorial Aplicada à Compressão de Sinais de Voz e Imagem*. Dissertação (Dissertação de Mestrado) — Universidade Federal da Paraíba - Campus II, Campina Grande, Brasil, Março 1998.

BOVIK, A.; HUANG, T.; MUNSON D., J. A generalization of median filtering using linear combinations of order statistics. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. 31, n. 6, p. 1342–1350, Dec 1983. ISSN 0096-3518.

BROADCASTING, D. V. *DVB Approves DVB-SH Specification - New specification addresses delivery of multimedia services to hybrid satellite/terrestrial mobile devices*. 2007.

BROOKS, M. Development of broadcast technologies for mobile TV. *The IEE Seminar on Broadcasting Spectrum: The Issues*.(Ref. No. 2005/11049), p. 16, June 2005. ISSN 0537-9989.

CHORIANOPOULOS, K. Personalized and mobile digital tv applications. *Multimedia Tools and Applications*, v. 36, n. 1-2, January 2008. ISSN 1380-7501 (Print) 1573-7721 (Online).

FACON, J. *Apostila de Processamento e Análise de Imagens*. <http://www.ppgia.pucpr.br/facon/MaterialGraduacao2005/ApostilaProcImagem.pdf>, Fevereiro 2002.

FARIA, G. *et al.* DVB-H: Digital broadcast services to handheld devices. *Proceedings of the IEEE*, v. 94, n. 1, p. 194–209, Jan. 2006. ISSN 0018-9219.

FECHINE, J. M. *Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística*. Tese (Tese de Doutorado) — Universidade Federal da Paraíba - UFPB - Campus II, Campina Grande - PB, Brasil, Dezembro 2000.

FONSECA, R. N. da. *Algoritmos para Avaliação da Qualidade de Vídeo em Sistemas de televisão Digital*. Dissertação (Dissertação de Mestrado) — Escola Politécnica da Universidade de São Paulo, São Paulo, Brasil, 2008.

GONÇALVES, G. A.; GARCIA, C. A. E. Processamento digital de imagens SAR/ERS-1 na caracterização de feições oceanográficas. *VIII Simpósio Brasileiro de Sensoriamento Remoto*, p. p. 769–773., abril 1996.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001. ISBN 0201180758.

GROUP, D. T. T. A. *Television on a handheld receiver, broadcasting with DVB-H*. 2005. Geneva, Switzerland.

HANDS, D. A basic multimedia quality model. *IEEE Transactions on Multimedia*, v. 6, n. 6, p. 806–816, Dec. 2004. ISSN 1520-9210.

HOLMA, H.; TOSKALA, A. *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. 3a. ed. Londres: John Wiley & Sons, 2004.

ITU-R. *ITU-R BT.500-11 - Methodology for the subjective assessment of the quality of television pictures*. 2002.

ITU-T. *ITU-T Recommendation P.910, Subjective video quality assessment methods for multimedia applications*. September 1999.

ITU-T. *ITU-T Video coding for low bit rate communication*. November 2000.

ITU-T, J. V. T. of; 1, I. J. *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 ISO/IEC 14496-10 AVC)*. [S.l.], 2003.

ITU-T, J. V. T. of; 1, I. J. *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 ISO/IEC 14496-10 AVC) – Fidelity Range Extensions documents JVT-L047*. [S.l.], Julho 2003.

KIM, C. S. *et al.* Measuring video quality on full scalability of H.264/AVC scalable video coding. *IEICE TRANSACTIONS on Communications*, E91-B, n. 5, p. 1269–1278, May 2008. ISSN 1745-1345.

KNOCHE, H.; MCCARTHY, J. D. Mobile users' needs and expectations of future multimedia services. *Proceedings of the WWR12*, 2004.

KNOCHE, H.; MCCARTHY, J. D. Design requirements for mobile tv. *MobileHCI'05*, Setembro 19-22 2005.

KNOCHE, H.; MCCARTHY, J. D.; SASSE, M. A. Can small be beautiful? Assessing image resolution requirements for mobile TV. ACM, New York, NY, USA, p. 829–838, 2005.

KNOCHE, H.; MCCARTHY, J. D.; SASSE, M. A. How low can you go? The effect of low resolutions on shot types in mobile TV. ACM, London, UK, p. 145–166, 2008.

KNOCHE, H. O.; SASSE, M. A. Breaking the news on mobile TV: User requirements of a popular mobile content. *Proc. SPIE*, Vol. 6074, 2006.

KUMAR, A. *Mobile TV: DVB-H, DMB, 3G Systems and Rich Media Applications*. [S.l.]: Focal Press Media technology Professional, 2007.

LABGIS. *Apostila de Processamento de Imagens – Laboratório de Geoprocessamento do Departamento de Geologia Aplicada da Faculdade de Geologia*. 2008. Acessado em 17 de Julho.

- LEE, Y.-R. *et al.* Low-complexity DCT-domain video transcoders for arbitrary-size downsampling. *2004 IEEE 6th Workshop on Multimedia Signal Processing*, p. 31–34, Sept.-1 Oct. 2004.
- LUKAC, R. *et al.* Generalized adaptive vector sigma filters. *International Conference on Multimedia and Expo. ICME '03.*, v. 1, p. I-537–40 vol.1, July 2003.
- MATSUMURA, K. *et al.* A practical experiment on mobile data broadcasting. *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, p. 337–342, 3-6 Jan 2005.
- MERRIT, L. *et al.* X264. march 2007. <http://www.x264.nl>.
- MITRA, S. K. *Digital Signal Processing: A Computer Based Approach*. 3a. ed. Santa Barbara: MacGraw-Hill, 2006.
- MOREIRA., D. *Oi anuncia TV para celular e promete IPTV para o próximo trimestre*. <http://sbtvd.cpqd.com.br/>, Acessado em 8 de Novembro de 2007 2007.
- MUNTEAN, G.; PERRY, P.; MURPHY, L. Objective and Subjective Evaluation of QOAS Video Streaming over Broadband Networks. *IEEE Electronic Transactions on Network and Service Management*, v. 2, n. 1, p. 19–28, Novembro 2005.
- OLIVEIRA, J. F. F. de. *Uma Análise do Canal Móvel do Sistema Brasileiro de Televisão Digital*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Campina Grande, Campina Grande, Brasil, Dezembro 2008.
- PAIVA, M. C. de; FERREIRA, J. S.; MENDES, L. L. Soluções de simulação e implementação em fpga de um modulador ISDB-TB. *XXVI Simpósio Brasileiro de Telecomunicações - SBrT 2008*, Sept. 2008.
- PITAS, I.; VENETSANOPOULOS, A. Nonlinear mean filters in image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. 34, n. 3, p. 573–584, Jun 1986. ISSN 0096-3518.
- POYNTON, C. *Digital Video and HDTV: Algorithms and interfaces*. San Francisco: Morgan Kaufmann Publishers, 2003.
- QUEIROZ, R. L. *Elementos básicos de transcodificação de vídeo para o Sistema Brasileiro de Televisão Digital (SBTVD) – Relatório Técnico*. Campina Grande, 2005.
- REGIS, C. D. M. *et al.* Transcodificação de vídeo digital para receptores portáteis. *IV SEGeT - Simpósio de Excelência em Gestão e Tecnologia*, 2007.
- REHME, J. F. *Avaliação da Qualidade de Vídeo Trafegando sobre Redes IP*. Dissertação (Dissertação de Mestrado) — Universidade Tecnológica Federal do Paraná, Curitiba, Brasil, Fevereiro 2007.

-
- RICHARDSON, I. E. G. *H.264 / MPEG-4 Part 10 White Paper – Reconstruction Filter*. Abril 2003.
- ROCHA, J. S. *Avaliação de Desempenho da Camada Física de Redes do Sistema de Telecomunicações Móveis Universal*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Campina Grande, Campina Grande, Brasil, Maio 2008.
- ROCHA, J. S. *et al.* Avaliação da transmissão de televisão digital móvel em redes UMTS. XXVI *Simpósio Brasileiro de Telecomunicações – SBrT’08*, Setembro 2008.
- ROGER, L. L. B. *Apostila de Estatística Geral e Aplicada*. <http://www.ceset.unicamp.br/webdidat/matdidat.php?cod=TT%20203&nome=Leonardo+Lorenzo+Bravo+Roger>: UNICAMP, 2007.
- SAMSUNG. *Samsung lança primeiro celular com suporte a TV digital móvel e 3G*. 2008. <Http://idgnow.uol.com.br/telecom/2008/03/04/samsung-lanca-primeiro-celular-com-suporte-tv-digital-movel-e-3g/>, Acessado em 22/10/08.
- SCHWARTZ, W. R. *Segmentação de Imagens Baseadas em Dependência Espacial Utilizando Campo de Aleatório de de Markov Associado com Características e Texturas*. Dissertação (Dissertação de Mestrado) — Universidade Federal do Paraná, Curitiba, Brasil, 1995.
- SHANABLEH, T.; GHANBARI, M. Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats. *IEEE Transactions on Multimedia*, v. 2, n. 2, p. 101–110, Jun. 2000. ISSN 1520-9210.
- SHIGUTI, W. A.; SHIGUTI, V. da S. C. *Apostila de Estatística*. Brasília: [s.n.], 2006.
- SHMILDT, E. R. *Apostila de BIOESTATÍSTICA: ESTATÍSTICA DESCRITIVA – Unidade II*. 2007.
- STEVERSON, W. J. *Estatística Aplicada à Administração*. São Paulo: Editora Habra, 1986.
- TAKADA, M.; SAITO, M. Transmission system for ISDB-T. *Proceedings of the IEEE*, v. 94, n. 1, p. 251–256, Jan. 2006. ISSN 0018-9219.
- TRACE, W. *YUV Video Sequences*. Abril 2008.
- VAIDYANATHAN, P. P. *Multirate systems and filter banks*. New Jersey: Prentice Hall/Englewood Cliffs, 1993.
- VETRO, A.; CHRISTOPOULOS, C.; SUN, H. Video transcoding architectures and techniques: an overview. *Signal Processing Magazine, IEEE*, v. 20, n. 2, p. 18–29, Mar. 2003. ISSN 1053-5888.

-
- VRANJES, M.; RIMAC-DRLJE, S.; ZAGAR, D. Objective video quality metrics. *49th International Symposium ELMAR, 2007*, p. 45–49, Sept. 2007. ISSN 1334-2630.
- WANG, L. L. Z.; BOVIK, A. C. Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication*, v. 19, n. 2, p. 121–132, february 2004.
- WANG, Z.; BOVIK, A. A universal image quality index. *Signal Processing Letters, IEEE*, v. 9, n. 3, p. 81–84, Mar 2002. ISSN 1558-2361.
- WANG, Z. *et al.* Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, v. 13, n. 4, p. 600–612, 2004.
- WU, H.; RAO, K. *Digital Video Image Quality and Perceptual Coding*. Boca Raton, FL, USA: CRC Press Taylor & Francis Group, 2006. ISBN 0-8247-2777-0.
- XIN, J.; LIN, C.; SUN, M. Digital video transcoding. *Proceedings of the IEEE*, v. 93, n. 1, p. 84–97, Jan. 2005. ISSN 0018-9219.
- XIN, J. *et al.* An HDTV-to-SDTV spatial transcoder. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 12, n. 11, p. 998–1008, Nov. 2002. ISSN 1051-8215.
- XU, Y. *et al.* MV-based adaptive transcoding technique for reduced spatial resolution. *IEEE International Symposium on Consumer Electronics*, p. 324–327, 1-3, 2004.
- YIN, L. *et al.* Weighted median filters: a tutorial. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v. 43, n. 3, p. 157–192, Mar 1996. ISSN 1057-7130.
- YIN, P.; WU, M.; LIU, B. Video transcoding by reducing spatial resolution. *International Conference on Image Processing*, v. 1, p. 972–975, 2000.