

MINISTÉRIO DA DEFESA  
EXÉRCITO BRASILEIRO  
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA  
INSTITUTO MILITAR DE ENGENHARIA  
CURSO DE MESTRADO EM ENGENHARIA ELÉTRICA

DIOGO DIAS MUNHOZ

CODIFICAÇÃO DE SEQUÊNCIAS ESTÉREO USANDO O  
CODIFICADOR H.264

Rio de Janeiro  
2006

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**INSTITUTO MILITAR DE ENGENHARIA**

**DIOGO DIAS MUNHOZ**

**CODIFICAÇÃO DE SEQUÊNCIAS ESTÉREO USANDO O  
CODIFICADOR H.264**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Engenharia Elétrica do Instituto Militar de Engenharia, como requisito parcial para obtenção do título de Mestre em Ciências em Engenharia Elétrica.

Orientador: Prof. Carla Liberal Pagliari - Ph.D.

Co-orientador: Prof. Marcelo de Mello Perez - Ph.D.

Rio de Janeiro  
2006

c2006

INSTITUTO MILITAR DE ENGENHARIA  
Praça General Tibúrcio, 80-Praia Vermelha  
Rio de Janeiro-RJ CEP 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

M966 Munhoz, Diogo Dias  
Codificação de Seqüências Estéreo usando o Codificador H.264, Diogo Dias Munhoz.  
– Rio de Janeiro: Instituto Militar de Engenharia, 2006.  
125 p.:il, graf., tab.

Dissertação: (mestrado) – Instituto Militar de Engenharia, Rio de Janeiro, 2006.

1. Processamento de imagens. 2. Codificação de vídeo estéreo. 3. Codificador H.264. 4. Graph Cut. I. Título. II. Instituto Militar de Engenharia.

CDD 621.367

**INSTITUTO MILITAR DE ENGENHARIA**

**DIOGO DIAS MUNHOZ**

**CODIFICAÇÃO DE SEQUÊNCIAS ESTÉREO USANDO O  
CODIFICADOR H.264**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Engenharia Elétrica do Instituto Militar de Engenharia, como requisito parcial para obtenção do título de Mestre em Ciências em Engenharia Elétrica.

Orientador: Prof. Carla Liberal Pagliari - Ph.D.

Co-orientador: Prof. Marcelo de Mello Perez - Ph.D.

Aprovada em 09 de Abril de 2006 pela seguinte Banca Examinadora:

---

Prof. Carla Liberal Pagliari - Ph.D. - IME - Presidente

---

Prof. Marcelo de Mello Perez - Ph.D. - IME

---

Prof. Ernesto Leite Pinto - D.C. - IME

---

Prof. Eduardo Antônio Barros da Silva - Ph.D. - COPPE/UFRJ

Rio de Janeiro  
2006

Aos meus pais, Diogo Munhoz Fernandez e Alaíde  
Dias Munhoz, por tudo que representam na minha  
vida.

## AGRADECIMENTOS

A minha orientadora, Prof. Carla Liberal Pagliari. - Ph.D., pela atenção e profissionalismo com que acompanhou a realização deste trabalho e, sobretudo, pela grande amizade e incentivo que muito contribuíram para o desenvolvimento do meu trabalho.

Ao meu Co-orientador, Prof. Marcelo de Mello Perez - Ph.D. pela colaboração incondicional na leitura do manuscrito original e pelas valiosas sugestões.

A todos os professores e funcionários do Departamento de Engenharia Elétrica do Instituto Militar de Engenharia que, de alguma forma, contribuíram para a realização deste trabalho.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Aos meus amigos e colegas de pesquisa, pelo apoio e pelo convívio amigável. Em especial a Alessandra (Alessandreta), Roberto (Robertoca) e Paúl (Paulete) pelo apoio e pelas brincadeiras desde o começo do curso. E também aos amigos Caíque (Cabeça), Michele (Mimi), Arthur (Giga-Boy), Leo (Gnomo), Leo (Capitão), João (Janja Jones), Gleyson (Super-Man), Catia (Catita), Paulo (Fanta Uva), Loss (El Migué), Ricardo Toscano, Márcio Toscano, Giovanna, Itacy, Adriana, Glayce, Diogo, Torres, Paiva, Nicodemos, Bandeira e Sabrina.

Ao meu irmão Alexandre, que sempre esteve disponível para me ajudar e incentivar em todos os momentos.

Aos meus familiares, pela paciência, apoio e orações.

Ao meu amigo Rafael Pereira Baptista (Bambi), pela disponibilidade e amizade nesse período.

A Fernanda Guedes (Pinpin), pelos lanches, força, paz de espírito, amor e carinho.

“Nunca ande pelo caminho traçado, pois ele conduz  
somente até onde os outros foram.”  
Alexandre Graham Bell.

## SUMÁRIO

LISTA DE ILUSTRAÇÕES .....	9
LISTA DE TABELAS .....	12
LISTA DE ABREVIATURAS E SÍMBOLOS .....	14
<b>1 INTRODUÇÃO .....</b>	<b>18</b>
1.1 Introdução .....	18
1.2 Objetivo da Dissertação .....	20
1.3 Trabalhos da Dissertação .....	21
1.4 Organização .....	21
<b>2 ESTEREOSCOPIA .....</b>	<b>23</b>
2.1 Introdução .....	23
2.2 Estereoscopia .....	24
2.3 Métodos de Estimação de Disparidades .....	26
2.3.1 Problemas na Estimação de Disparidades .....	30
2.3.2 Estimador de Disparidades Usando o Método Graph Cut .....	31
2.4 Monitores Estereoscópicos. ....	35
2.5 Resumo .....	36
<b>3 CODIFICAÇÃO DE VÍDEO ESTÉREO .....</b>	<b>37</b>
3.1 Introdução .....	37
3.2 Codificação de Imagem .....	37
3.3 Codificação de Vídeo Monocular .....	40
3.4 O Codificador H.264/AVC .....	45
3.4.1 Estrutura do Codificador H.264/AVC .....	46
3.4.2 Camada de Codificação de Vídeo (CCV) .....	46
3.5 Codificação de Vídeo Estéreo .....	54
3.6 Resumo .....	61
<b>4 TESTE DO DESEMPENHO DOS MÉTODOS .....</b>	<b>62</b>
4.1 Introdução .....	62

4.2	Ambiente de Teste .....	63
4.2.1	Seqüências Utilizadas .....	64
4.2.2	Medidas de Avaliação .....	65
4.3	Ganho de Codificação Utilizando o Método “Entrelaçado” .....	66
4.4	Comparação entre os Métodos - Taxa de Bits Constante .....	71
4.4.1	Resultados da Seqüência MAN com o Perfil <i>Main</i> e <i>High</i> . .....	72
4.4.2	Resultados da Seqüência SAXO com o Perfil <i>Main</i> e <i>High</i> . .....	78
4.4.3	Resultados da Seqüência INDOOR com o Perfil <i>Main</i> e <i>High</i> . .....	80
4.4.4	Resultados da Seqüência OUTDOOR com o Perfil <i>Main</i> e <i>High</i> . .....	84
4.5	Comparação entre os Métodos - Taxa de Bits Variável .....	88
4.5.1	Resultados - Seqüência MAN. ....	88
4.5.2	Resultados - Seqüência SAXO. ....	90
4.5.3	Resultados - Seqüência INDOOR. ....	93
4.5.4	Resultados - Seqüência OUTDOOR. ....	95
4.6	Comparação entre os Métodos - Método “Binocular” .....	97
4.6.1	Resultados - Seqüência MAN. ....	99
4.6.2	Resultados - Seqüência SAXO. ....	101
4.6.3	Resultados - Seqüência INDOOR. ....	102
4.6.4	Resultados - Seqüência OUTDOOR. ....	106
4.6.5	Resultados - Seqüência RANDOT. ....	108
4.7	Resumo .....	110
<b>5</b>	<b>CONCLUSÕES E COMENTÁRIOS FINAIS</b> .....	<b>112</b>
5.1	Conclusões .....	112
5.2	Trabalhos Futuros .....	113
<b>6</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>115</b>
<b>7</b>	<b>APÊNDICES</b> .....	<b>122</b>
7.1	APÊNDICE 1: Imagens Originais .....	123

## LISTA DE ILUSTRAÇÕES

FIG.2.1	Projeção dos pontos $P$ e $Q$ . (FAUGERAS, 1993a; GONZALEZ, 1992). . . . .	24
FIG.2.2	SVH - Visão Estéreo. . . . .	25
FIG.2.3	Geometria básica do sistema binocular (PAGLIARI, 1999). . . . .	27
FIG.2.4	Problema de correspondência. . . . .	28
FIG.2.5	Problema de Oclusão. . . . .	32
FIG.2.6	Projeção Graph Cut. . . . .	33
FIG.3.1	Estrutura de um codificador genérico. . . . .	38
FIG.3.2	DC, Bordas e Altas Frequências - DCT. (PAGLIARI, 1999). . . . .	38
FIG.3.3	Imagem de teste Lena 512x512 . . . . .	41
FIG.3.4	Formação de uma Seqüência Monocular . . . . .	42
FIG.3.5	Quadro dividido em macroblocos. (SCHAFER, 2003; WIEGAND, 2003). . . . .	42
FIG.3.6	Seqüência MAN - Estimação de Movimento. . . . .	44
FIG.3.7	GOP com quadros I, P e B. $M=3$ e $N=12$ . . . . .	45
FIG.3.8	Diagrama do codificador H.264/AVC (SCHAFER, 2003; WIEGAND, 2003) . . . . .	47
FIG.3.9	Codificação Progressiva - Modos de predição (WANG, 2002a). . . . .	50
FIG.3.10	Codificação por Campo - Modos de predição (WANG, 2002a). . . . .	50
FIG.3.11	Formação de um Super macrobloco (WANG, 2002a). . . . .	51
FIG.3.12	Codificação por Campo - Modos de predição super macrobloco (WANG, 2002a). . . . .	51
FIG.3.13	Modos de predição Intra para blocos 4x4 e macroblocos 16x16 de luminância. . . . .	53
FIG.3.14	Métodos de codificação <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. . . . .	58
FIG.3.15	Estimador e compensador de disparidades. . . . .	59
FIG.3.16	Método de codificação estéreo com estimador de disparidades. . . . .	60
FIG.4.1	Seqüências de vídeo estéreo “Concatenadas” . . . . .	68
FIG.4.2	Seqüências de vídeo estéreo “Concatenada” e “Entrelaçada”. . . . .	69
FIG.4.3	Comparação do Método “Entrelaçado”. Seqüências estéreo MAN. . . . .	70

FIG.4.4	Comparação do Método “Entrelaçado”. Seqüências estéreo SAXO. ....	70
FIG.4.5	Comparação PQ - Seqüência MAN esquerda <i>Simulcast</i> - Perfil <i>Main</i> e <i>High</i> . ....	74
FIG.4.6	Comparação do Método “Entrelaçado” com/sem PAFF. GoP I PQ23 - Perfil <i>Main</i> - MAN. ....	75
FIG.4.7	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado” sem PAFF. Seqüência MAN - GoP I PQ23 - Perfil <i>Main</i> . ....	76
FIG.4.8	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado” sem PAFF. Seqüência SAXO - GoP I PQ39 - Perfil <i>Main</i> ....	79
FIG.4.9	Comparação do Método “Entrelaçado” com/sem PAFF. GoP I PQ23 - Perfil <i>Main</i> - INDOOR ....	81
FIG.4.10	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado” sem PAFF. Seqüência INDOOR - GoP I PQ23 - Perfil <i>Main</i> . ....	82
FIG.4.11	Comparação do Método “Entrelaçado” com/sem PAFF. GoP I PQ23 - Perfil <i>Main</i> - OUTDOOR ....	85
FIG.4.12	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado” sem PAFF. Seqüência OUTDOOR - GoP I PQ23 - Perfil <i>Main</i> . ....	86
FIG.4.13	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil <i>MainxHigh</i> - MAN. ....	89
FIG.4.14	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil <i>MainxHigh</i> - SAXO. ....	92
FIG.4.15	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil <i>MainxHigh</i> - INDOOR. ....	94
FIG.4.16	Comparação dos métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil <i>MainxHigh</i> - OUTDOOR ....	97
FIG.4.17	Comparação dos métodos <i>Simulcast</i> e “Binocular”. - Seqüência MAN - Perfil <i>Main</i> . ....	100
FIG.4.18	Comparação dos métodos <i>Simulcast</i> e “Binocular”. - Seqüência SAXO - Perfil <i>Main</i> ....	103
FIG.4.19	Comparação dos métodos <i>Simulcast</i> e “Binocular”. - Seqüência INDOOR - Perfil <i>Main</i> ....	105
FIG.4.20	Comparação dos métodos <i>Simulcast</i> e “Binocular”. - Seqüência OUTDOOR - Perfil <i>Main</i> ....	107

FIG.4.21	Comparação dos métodos <i>Simulcast</i> , “Concatenado”, “Entrelaçado” e “Binocular”. - Seqüência RANDOT - Perfil <i>Main</i> .....	109
FIG.7.1	Imagem de Teste Lena (512x512). .....	123
FIG.7.2	Par estéreo - RANDOT (256x256). .....	123
FIG.7.3	Pares estéreo - MAN (384x384) e SAXO (352x288). .....	124
FIG.7.4	Pares estéreo - INDOOR e OUTDOOR (352x288). .....	125

## LISTA DE TABELAS

TAB.4.1	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP II. TAXA FIXA - MAN .....	78
TAB.4.2	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP II. PSNR FIXO - MAN .....	78
TAB.4.3	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> . GoP II. TAXA FIXA - SAXO .....	80
TAB.4.4	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> . GoP II. PSNR FIXO - SAXO .....	81
TAB.4.5	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP I e II. TAXA FIXA - INDOOR .....	84
TAB.4.6	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP I e II. PSNR FIXO - INDOOR .....	84
TAB.4.7	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>High</i> . GoP I. TAXA FIXA - OUTDOOR .....	87
TAB.4.8	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>High</i> . GoP I. PSNR FIXO - OUTDOOR .....	87
TAB.4.9	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP II. - MAN .....	90
TAB.4.10	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP III. - MAN .....	91
TAB.4.11	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP II. - SAXO .....	93
TAB.4.12	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP III. - SAXO .....	93
TAB.4.13	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP II. - INDOOR .....	95
TAB.4.14	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP III. - INDOOR .....	96
TAB.4.15	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”. Perfil <i>Main</i> e <i>High</i> . GoP II. - OUTDOOR .....	98
TAB.4.16	Comparação - Métodos <i>Simulcast</i> , “Concatenado” e “Entrelaçado”.	

Perfil *Main* e *High*. GoP III. - OUTDOOR ..... 98

## LISTA DE ABREVIATURAS E SÍMBOLOS

### ABREVIATURAS

2D	-	Bi-Dimensional
3D	-	Tri-Dimensional
AVC	-	Advanced Video Coding (Codificação de Vídeo Avançada)
CCV	-	Camada de Codificação de Vídeo
CAS	-	Camada de Abstração de Sistema
CD	-	Compensador de Disparidades
CAVLC	-	Context-Adaptive Variable-Length Coding
CABAC	-	Context-Adaptive Binary Arithmetic Coding
DCT	-	Discrete Cosine Transform (Transformada Cosseno Discreta)
DPCM	-	Differential Pulse Code Modulation (Modulação por Código de Pulso Diferencial)
EMQ	-	Erro Médio Quadrático
ED	-	Estimador de Disparidades
FRExt	-	Fidelity Range Extensions
GoP	-	Group of Pictures (Grupo de imagens)
IMDE	-	Interpolated Motion and Disparity Estimation
JVT	-	Joint Video Team
MPEG	-	Moving Pictures Expert Group
MBAFF	-	MacroBlock Adaptive Frame-Field
MD	-	Mapa de Disparidades (esquerdo - direito)
MDDECOD	-	Mapa de Disparidades Decodificado (esquerdo - direito)
PSNR	-	Peak Signal-to-Noise Rate (Relação Sinal Ruído de Pico)
PAFF	-	Picture Adaptive Frame-Field
PQ	-	Parâmetro de Quantização
RCD	-	Resíduo do Compensador de Disparidades
RCDCOD	-	Resíduo do Compensador de Disparidades Codificado
SEI	-	Supplemental Enhancement Information
SVH	-	Sistema Visual Humano
SDA	-	Somatório das Diferenças Absolutas

SDQ	-	Somatório das Diferenças ao Quadrado
SFEI-BLCF	-	Stereo Frame Estimation Interpolation-Bayesian Least Cost Function
TBV	-	Taxa de Bits Variável
TBC	-	Taxa de Bits Constante
VCEG	-	Video Coding Expert Group
VE	-	Vista Esquerda
VECOD	-	Vista Esquerda Codificada
VEDECOD	-	Vista Esquerda Decodificada
VD	-	Vista Direita
$VD_{rec}$	-	Vista Direita Reconstruída

## RESUMO

Esta tese avalia o desempenho de quatro métodos de codificação de seqüências estéreo usando o codificador H.264/AVC (Advanced Video Coding). O primeiro método codifica cada seqüência estéreo como duas seqüências monoculares independentes explorando somente as redundâncias espaciais e temporais das mesmas, sem explorar as redundâncias binoculares presentes em cada par estéreo das seqüências. Este método aumenta a complexidade no tratamento dos dois trens de bits resultantes do processo de codificação, pois o mesmo necessita de sincronização na decodificação das seqüências no lado do decodificador. O segundo e o terceiro métodos, combinam cada quadro esquerdo e direito das seqüências estéreo em um único quadro contendo as duas vistas gerando assim uma única seqüência de quadros monoculares a ser codificada. Eles diminuem a complexidade da manipulação do trem de bits resultante do processo de codificação. Estes dois métodos usam a mensagem "Supplemental Enhancement Information (SEI)", e o decodificador pode sincronizar as vistas, bem como identificar a existência de um trem de bits estéreo. Assim como o método anterior, eles exploram somente as redundâncias espaciais e temporais das seqüências estéreo. O quarto método utiliza um estimador de disparidades no processo de codificação das seqüências estéreo. Esse método explora além das redundâncias espaciais e temporais, as redundâncias binoculares contidas em cada par estéreo. No entanto, o método acarreta um aumento na complexidade no manuseio do trem de bits resultante do processo de codificação pois gera três trens de bits, sendo então necessário sincronizar a decodificação das seqüências no decodificador. Outra desvantagem do quarto método é a necessidade de estimação e transmissão do mapa de disparidades no processo de codificação.

O objetivo de se avaliar o quarto método, apesar da geração de três trens de bits, é avaliar a contribuição do estimador de disparidades no processo de codificação. Este ponto é justificável uma vez que os estimadores de movimento projetados para codificação de vídeo monocular podem não ser tão eficientes quando aplicados a vídeo estéreo devido às distorções perspectivas existentes nos pares estéreo. Entretanto, o codificador H.264/AVC que oferece suporte a múltiplos quadros de referência, diferentes tamanhos de blocos, estimação com precisão de sub-pixel, predição-intra e diferentes ferramentas de vídeo entrelaçado mostrou-se eficiente na codificação de vídeos estéreo.

No entanto, ainda há a necessidade de se investigar o desempenho dos estimadores de movimento dos codificadores de vídeo monoculares que sejam eficientes frente à distorções perspectivas mais acentuadas. O objetivo é recomendar métodos capazes de obter boa taxa de compressão mantendo a compatibilidade com os decodificadores monoculares. Seqüências estéreo com diferentes graus de movimento e textura são avaliadas nesta tese.

## ABSTRACT

This thesis evaluates the performance of four stereo video coding methods using the H.264/AVC standard. The first one codes each stereo sequence separately as two independent monocular sequences exploring only their spatial and temporal redundancies, without exploring the binocular redundancies present in each stereo pair of the sequences. As two bitstreams are being transmitted, the decoding process will have to synchronize the decoded views at the decoder side. The second and third methods, combine each left and right views of the stereo sequence in only one composite-frame. They reduce the complexity in the handle of the resultant bitstream of the coding process. As these methods use the stereo video Supplemental Enhancement Information (SEI) message, the decoder can synchronize the views, as well as identify a coded stereo video bitstream. They only exploit the spatial and temporal redundancies of the stereo sequences. The fourth method uses a disparity estimator in the coding process of the stereo sequences. This method exploits the spatial, temporal and binocular redundancies of each stereo pair. Nevertheless, it increases the complexity in the handling of the resultant bitstream as the method produces three bitstreams. As three bitstreams are being transmitted, the decoding process will have to synchronize the decoded views at the decoder side. Another drawback present in the fourth method is the estimation and transmission of the disparity map in the coding process.

The goal of the fourth method, despite the generation of three different bitstreams, is to evaluate the contribution of the disparity estimator to the coding process. As the motion estimators are designed to cope with monocular sequences, they can be not very efficient when dealing with perspective distorted cues. However, the H.264/AVC codec that has support of multiple references, block sizes, subpel accuracy, intra prediction, different interlace tools, etc has efficiently coded stereo video.

However, for the best stereo video coding performance, the motion estimation methods of monocular coders have to be efficient when handling more accentuate perspective distortions. The idea is to recommend methods capable to get good rate of compression keeping the compatibility with the monocular codecs. Stereo sequences with different degrees of movement and texture are evaluated in this thesis.

# 1 INTRODUÇÃO

## 1.1 INTRODUÇÃO

Os sistemas atuais de captura e transmissão de vídeo geram e transmitem seqüências monoculares, não explorando a informação de profundidade da cena original existente em sistemas binoculares.

Sistemas monoculares tentam extrair o máximo possível de informações que a cena original possa conter. Entretanto, no processo de captura da mesma, ocorre uma projeção da cena original 3D (Tri-Dimensional) sobre o plano de câmera 2D (Bi-Dimensional), sendo a informação de profundidade da cena original perdida. Este processo é não reversível, pois diferentes objetos em uma cena 3D podem ter suas projeções no plano de câmera 2D coincidentes.

Nos sistemas estereoscópicos esta informação é preservada, mas não o contrário. Nestes sistemas, a captura da cena original 3D é feita através de duas ou mais câmeras. A cena original 3D é então projetada no mínimo em dois planos de imagem, cada um associado a uma câmera, e uma vez que seja estabelecida a relação de correspondência entre os pontos homólogos dos planos de imagem das câmeras esquerda e direita, a informação de profundidade da cena original 3D poderá ser obtida por triangulação e pela geometria estéreo. O desenvolvimento de um método que estabeleça a correspondência entre os pontos homólogos dos planos de imagem da câmera esquerda e direita tem sido um desafio em sistemas de vídeos binoculares. Uma solução definitiva ainda está longe de ser elaborada, e existe campo suficiente para novas contribuições.

Diversas aplicações, nas áreas de entretenimento, medicina e operações remotas (HASKELL, 1997), (KOPERNIK, 1993), (MOTOKI, 1995), (ROSE, 1993), entre outras, têm-se beneficiado com a utilização de imagens com maior realismo (3D), dada pela sensação de percepção de profundidade da cena.

Por isso, a codificação de vídeo estéreo constitui-se, atualmente, num campo de grande importância nas áreas de processamentos de imagens e visão computacional, tendo um grande avanço devido à enorme importância da visão, que constitui a principal fonte de

informação.

Os codificadores de vídeo existentes no mercado foram projetados para codificar seqüências monoculares. O codificador H.264/AVC (Advanced Video Coding) (ITU-T, 2005) apresenta o melhor desempenho (SCHAFER, 2003), (WIEGAND, 2003), (HONG, 2002), (RICHARDSON, 2002) dentre os codificadores atuais na codificação de seqüências monoculares.

Seqüências monoculares apresentam redundâncias espaciais e temporais que são exploradas pelos codificadores. Já uma seqüência binocular apresenta redundâncias binoculares, a cada par de quadros estéreo, além das redundâncias anteriores. Seqüências binoculares apresentam alguns problemas adicionais como o do casamento estéreo, que é a obtenção da correspondência de um ponto da cena original projetado no plano de imagem esquerdo com a sua projeção no plano de imagem direito. A posição destes pontos, nos planos de imagens esquerdo e direito, em geral apresentam uma diferença. Essa diferença é chamada de disparidade estéreo (DUDA, 1973). Uma outra observação é sobre a visualização destas seqüências. As mesmas necessitam de monitores especiais que utilizam técnicas que possibilitam a visualização das seqüências estéreo (WOJCIECH, 2004; DAVI, 2003; LIPTON, 1991).

Além do problema da estimação das disparidades e da visualização, temos o dobro da taxa de codificação na transmissão ou no armazenamento, comparada com uma seqüência monocular, quando a seqüência binocular é codificada simplesmente como duas seqüências monoculares.

Muitas soluções já foram apresentadas para estes tipos de problema. Porém, a busca por soluções cada vez mais robustas mantendo a qualidade, a um custo computacional aceitável, tem motivado a continuação de pesquisas na área para o desenvolvimento de novos métodos de codificação de vídeo estéreo (PAGLIARI, 1999).

Um codificador de vídeo estéreo é projetado de forma a explorar características da seqüência estéreo como redundâncias binoculares, além da exploração das redundâncias espaciais e temporais, sempre com o objetivo de representar a seqüência de entrada com uma taxa de bits inferior à inicial, em conjunto com o nível de qualidade requerido ou aceitável para a seqüência binocular reconstruída.

Neste trabalho, dois dos quatro métodos de codificação de vídeo estéreo avaliados e comparados foram propostos em (SUN, 2003). Estes métodos, por adaptarem uma seqüência estéreo em uma monocular, foram propostos com a finalidade de diminuir a

complexidade do manuseio do trem de bits no codificador/decodificador, pois, há a transmissão de um único trem de bits ao invés de dois, como no método *Simulcast* (TSENG, 1996). Em (SUN, 2003, 2005), os dois métodos são inseridos dentro do código do codificador H.264/AVC e as simulações foram feitas somente com taxas de bits variáveis, utilizando seqüências estéreo não referenciadas, com valores de disparidades baixos e um número grande de quadros de referência na estimação de movimento. Nesta dissertação, o código do codificador não foi alterado. As seqüências “concatenadas” e “entrelaçadas” foram montadas fora do codificador. As seqüências utilizadas são referenciadas e apresentam valores de disparidades baixos, médios e altos. As simulações foram realizadas com valores de quadros de referência baixos e altos, com três tipos de *Group of Pictures* (GoP) e com taxas de bits constantes e variáveis.

## 1.2 OBJETIVO DA DISSERTAÇÃO

O objetivo desta dissertação é estudar, avaliar e comparar o desempenho de quatro métodos de codificação estéreo. O primeiro método de codificação estéreo é denominado *Simulcast* (HASKELL, 1997; TSENG, 1996), o segundo e o terceiro estão presentes na proposta feita em (SUN, 2003), que combina em um mesmo quadro monocular os dois quadros correspondentes às vistas esquerda e direita de uma seqüência binocular. O último método explora as redundâncias binoculares da seqüência estéreo através de um estimador de disparidades de eficiência comprovada.

Desta forma, foram implementados e avaliados quatro métodos conforme listados abaixo:

- ***Simulcast***. Os quadros da seqüência esquerda e a da direita são codificados separadamente.
- **“Concatenado”**. Cada quadro da seqüência esquerda é concatenado ao da direita, formando um novo quadro, de largura duas vezes maior, que é codificado.
- **“Entrelaçado”**. Cada quadro da seqüência esquerda é entrelaçado ao da direita, formando um novo quadro de altura duas vezes maior, sendo a primeira linha do novo quadro a primeira linha do quadro esquerda, e a segunda linha do novo quadro a primeira linha do quadro direito, seguindo esta ordem até a última linha do novo quadro. Este quadro “entrelaçado” é codificado.

- **“Binocular”**. A seqüência esquerda é codificada diretamente como uma seqüência monocular. Cada quadro da esquerda, juntamente com o seu correspondente da direita são enviados ao estimador de disparidades, que por sua vez gera um mapa de disparidades. Uma combinação de cada quadro esquerdo, mapa de disparidades associado e resíduo correspondente gera o quadro direito reconstruído. Este processo envolve três fases distintas: codificação da seqüência esquerda, estimação dos mapas de disparidades e obtenção do resíduo.

### 1.3 TRABALHOS DA DISSERTAÇÃO

Nesta dissertação é feita uma avaliação do desempenho de quatro métodos de codificação de vídeo estéreo utilizando o codificador H.264/AVC. Com base em simulações, são apresentadas as melhores opções de transmissão de vídeo estéreo, entre os métodos. Como será visto, a utilização dos métodos dois (“Concatenado”) e três (“Entrelaçado”) são alternativas viáveis por apresentarem um ótimo ganho de compressão com uma boa qualidade e permitirem que se utilize um decodificador monocular na decodificação de seqüências binoculares.

Os testes indicam que o método “Entrelaçado” obtém um desempenho inferior para pares estéreo com um valor de disparidade elevado. Uma conclusão mais precisa não foi possível por disponibilizar para os testes, de uma única seqüência estéreo com o valor das disparidades elevado. É implementado e avaliado um método baseado na redundância binocular da seqüência estéreo, denominado nesta dissertação de método “Binocular”. Este método utiliza um estimador de disparidades usando o método graph cut.

Este método apresentou uma taxa de codificação superior aos outros métodos comparados. O aumento do desempenho deste método está associado à qualidade do mapa de disparidades gerado pelo estimador de disparidades. A título ilustrativo foi provada a eficiência do método através de uma seqüência artificial onde cada par estéreo possuía um mapa de disparidades com *Ground Truth* (disparidades conhecidas).

### 1.4 ORGANIZAÇÃO

Esta dissertação está organizada da seguinte forma: um estudo sobre estereoscopia,

estimadores de disparidades e o estimador de disparidades usando o método graph cut utilizado nesta dissertação, são apresentados no Capítulo 2. No Capítulo 3 são apresentados os conceitos básicos de vídeo monocular e binocular, a justificativa da utilização do codificador H.264/AVC e suas características, e uma breve descrição de codificadores de vídeo estéreo juntamente com a descrição dos quatro métodos utilizados nesta dissertação. No Capítulo 4 o ambiente de teste é descrito e são apresentados os resultados comparativos entre os quatro métodos com taxas de bits constantes e variáveis. A conclusão destes resultados e comentários finais são apresentados no Capítulo 5.

## 2 ESTEREOSCOPIA

Neste capítulo são apresentados conceitos básicos sobre estereoscopia, bem como alguns estimadores de disparidades juntamente com o estimador de disparidades usando o método graph cut utilizado nesta dissertação. Na Seção 2.1 é feita uma breve introdução sobre vídeo monocular e binocular. Conceitos importantes sobre estereoscopia são apresentados na Seção 2.2 e os principais métodos de estimação de disparidades estéreo, juntamente com o estimador de disparidades utilizado nesta dissertação são descritos na Seção 2.3. Na Seção 2.4 é feita uma breve introdução sobre monitores usados na exibição de vídeo 3D. Por fim na Seção 2.5, apresenta-se o resumo do capítulo.

### 2.1 INTRODUÇÃO

Atualmente, a maioria dos sistemas de processamento de imagens gera e transmite vídeo monocular. Sistemas monoculares apresentam uma característica onde objetos diferentes da cena do mundo real 3D podem obter a mesma projeção no plano de imagem da câmera 2D. Desta forma, a informação de profundidade dos objetos da cena do mundo real 3D é perdida, pois a cena do mundo real 3D é projetada no plano de imagem da câmera que é 2D.

A FIG. 2.1 ilustra o processo de captura da informação de dois pontos ( $P$  e  $Q$ ) de objetos diferentes da mesma cena do mundo real 3D em um sistema monocular formado por uma câmera, onde o ponto  $P$  encontra-se atrás do ponto  $Q$ . Na FIG. 2.1,  $P'$  e  $Q'$  são as projeções no plano de imagem dos pontos  $P$  e  $Q$ , sendo o ponto  $O$  o centro óptico da câmera e  $f$  a distância focal da referida câmera (FAUGERAS, 1993a; GONZALEZ, 1992). Observa-se que a projeção dos pontos  $P$  e  $Q$  no plano de imagem encontra-se na mesma coordenada  $(x, y)$ , sendo a informação de profundidade dos pontos  $P$  e  $Q$  perdida.

Em aplicações onde a informação de profundidade da cena não se faz necessária, estes sistemas são completamente satisfatórios.

Já os sistemas binoculares, surgiram da necessidade de recuperar a informação de profundidade da cena do mundo real 3D, beneficiando aplicações que fazem uso direto

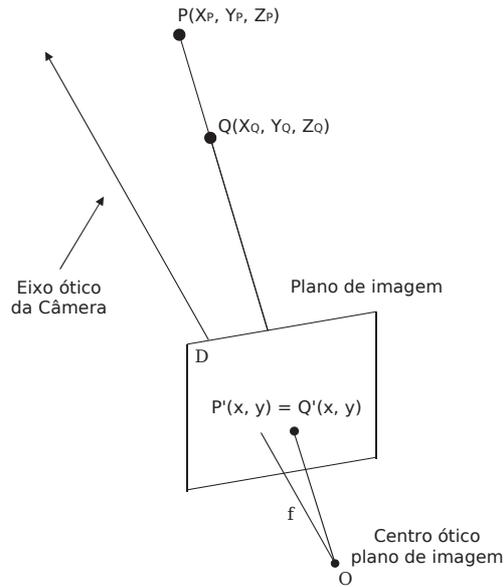


FIG. 2.1: Projeção dos pontos  $P$  e  $Q$ . (FAUGERAS, 1993a; GONZALEZ, 1992).

dessa informação, como sistemas de medida de distâncias, velocidades e acelerações, sistemas para interpretação de imagens e para visão computacional (HASKELL, 1997; KOPERNIK, 1993; MOTOKI, 1995; ROSE, 1993), uma vez que, as técnicas desenvolvidas para a extração dessa informação em sistemas monoculares mostraram resultados imprecisos e ambíguos (FAUGERAS, 1993a; PEREZ, 2003).

## 2.2 ESTEREOSCOPIA

No processo de captura da informação do objeto 3D da cena do mundo real através do SVH (Sistema Visual Humano) ou de um sistema formado de pelo menos duas câmeras, um ponto do objeto 3D se projeta nos planos retiniais esquerdo e direito ou nos planos de imagem da câmera esquerdo e direito, sendo este processo caracterizado por um mapeamento de um ponto do objeto 3D no plano de imagem da câmera (2D), onde a informação de profundidade do ponto do objeto 3D é perdida (FAUGERAS, 1993a; GONZALEZ, 1992). A informação capturada do ponto do objeto 3D da cena do mundo real projetada

nos planos retiniais esquerdo e direito ou planos de imagem corresponde à medida da luz refletida pelo ponto nestes planos. Esta informação capturada do ponto do objeto 3D é denominada de elementos de imagem (pixels), e representa o valor da intensidade da luz refletida pelo ponto nestes planos.

Tanto no SVH, quanto no sistema formado por pelo menos duas câmeras, a projeção do ponto do objeto 3D da cena do mundo real, nos planos de imagem esquerdo e direito, apresenta uma pequena diferença na sua posição. Esta diferença é denominada de disparidade estéreo (DUDA, 1973; SCHALKOFF, 1989; FAUGERAS, 1993a; TRUCCO, 1998). A disparidade estéreo é descrita em termos de suas componentes horizontais e verticais.

A habilidade do SVH ou de um sistema, composto de pelo menos duas câmeras, de extrair a informação de profundidade (3D) dos pontos do objeto da cena do mundo real 3D, provida pela disparidade estéreo é denominada de estereoscopia (HOWARD, 1995).

Na FIG. 2.2, é mostrada a reconstrução da imagem pelo cérebro através da diferença relativa (disparidade) da posição da projeção dos pontos da cena do mundo real 3D, nos planos retiniais esquerdo e direito.

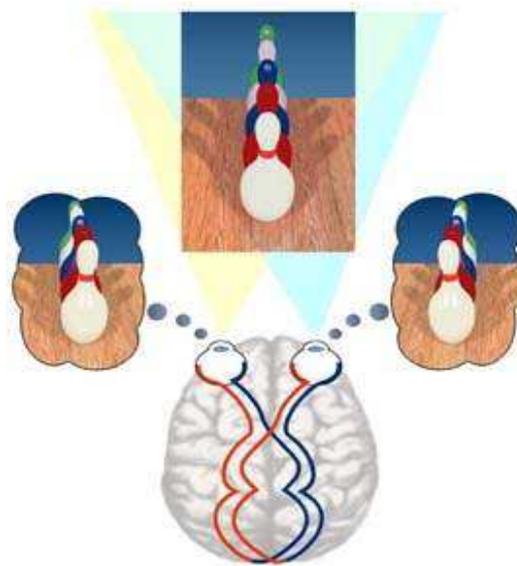


FIG. 2.2: SVH - Visão Estéreo.

Um problema presente em sistemas binoculares está no estabelecimento de pontos homólogos pertencentes aos dois planos de projeção ou planos de imagem do ponto do objeto da cena do mundo real projetado nos mesmos. Ou seja, determinar qual ponto no plano de imagem esquerdo corresponde ao ponto no plano de imagem direito (DUDA,

1973; SCHALKOFF, 1989; FAUGERAS, 1993a; TRUCCO, 1998). Para cada ponto projetado no plano de imagem esquerdo há somente um correspondente no plano de imagem direito (HOWARD, 1995). A solução deste problema de correspondência é muito importante para que se possa recuperar a informação de profundidade da cena do mundo real 3D. O grau de confiabilidade desta informação é diretamente proporcional à acurácia na solução do problema de correspondência. Este problema será mostrado com mais detalhes na Seção 2.3.

## 2.3 MÉTODOS DE ESTIMAÇÃO DE DISPARIDADES

Métodos de estimação de disparidade (Seção 2.2), também conhecidos como métodos de casamento estéreo ou correspondência estéreo têm o intuito de resolver o problema de correspondência presente em cada par de imagens estéreo. Estes métodos estabelecem a correspondência ou não, caso haja oclusão ou outra dificuldade no processo, da projeção do ponto  $P(X,Y,Z)$  da cena do mundo real (FIG. 2.3) no plano de imagem esquerdo  $(X_E, Y_E)$  com o seu correspondente no plano de imagem direito  $(X_D, Y_D)$ . A distância entre os dois centros ópticos representada (FIG. 2.3) pelo segmento de linha  $\overline{O_e O_d}$  é denominada de *Baseline* (FAUGERAS, 1993a; GONZALEZ, 1992). Os pontos  $C_e$  e  $C_d$  são denominados pontos principais e de fixação (FAUGERAS, 1993a; GONZALEZ, 1992). O plano formado pelos pontos  $O_e P O_d$  é denominado plano epipolar e a interseção deste plano com os planos de imagens das câmeras esquerda e direita gera as linhas epipolares. A interseção das linhas epipolares esquerda e direita com a *Baseline* geram os pontos epipoles esquerdo e direito (FAUGERAS, 1993a; GONZALEZ, 1992). As duas câmeras apresentam a mesma distância focal  $f$  representada pelos segmentos de linha  $\overline{O_e C_e}$  e  $\overline{O_d C_d}$ .

A disparidade da vista esquerda para a direita (Desq) é dada por:

$$Desq = (X_D - X_E, Y_D - Y_E) \quad (2.1)$$

e se a correspondência fosse do plano de imagem direita para o esquerdo à disparidade direita (Ddir) seria:

$$Ddir = (X_E - X_D, Y_E - Y_D) \quad (2.2)$$

É assumido que as câmeras possuem seus eixos ópticos paralelos e perpendiculares à *baseline*, separados por um deslocamento (*baseline*), com a mesma distância focal.

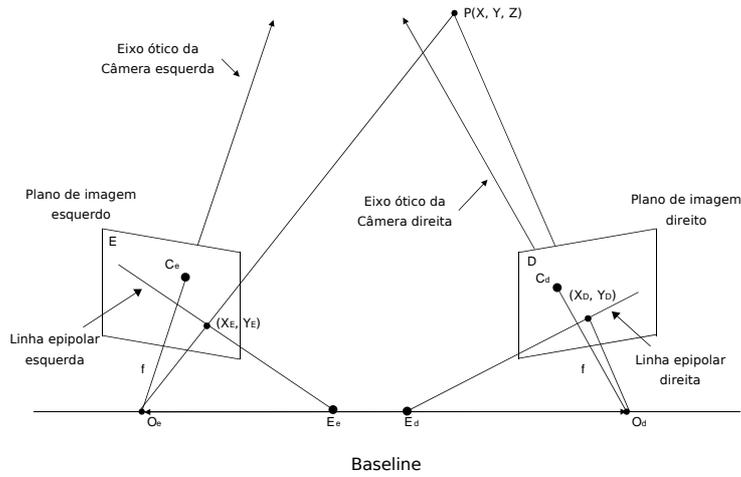


FIG. 2.3: Geometria básica do sistema binocular (PAGLIARI, 1999).

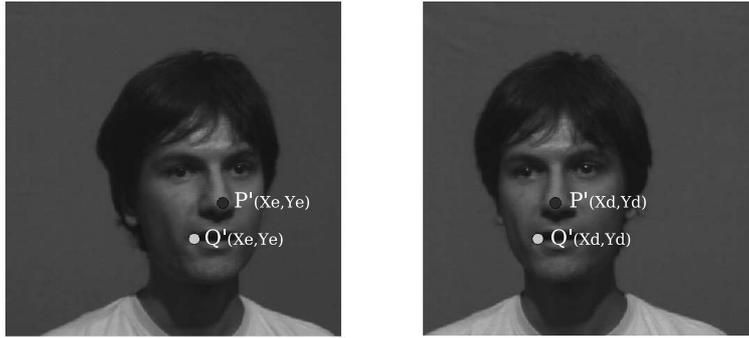
Também se admite que os pontos do objeto 3D que se projetam no plano de imagem esquerdo da câmera sobre suas respectivas linhas epipolares correspondem aos pontos que se projetam no plano de imagem direito da câmera sobre as linhas epipolares homólogas, tornando a busca dos pontos correspondentes de 2D para 1D (DUDA, 1973; SCHALKOFF, 1989; FAUGERAS, 1993a; TRUCCO, 1998). Desta forma, a disparidade é dada somente na componente horizontal e as linhas epipolares coincidem às linhas de varredura das imagens (DUDA, 1973; SCHALKOFF, 1989; FAUGERAS, 1993a; TRUCCO, 1998).

Com isso, as disparidades nas vistas são simétricas (PEREZ, 2003). Logo:

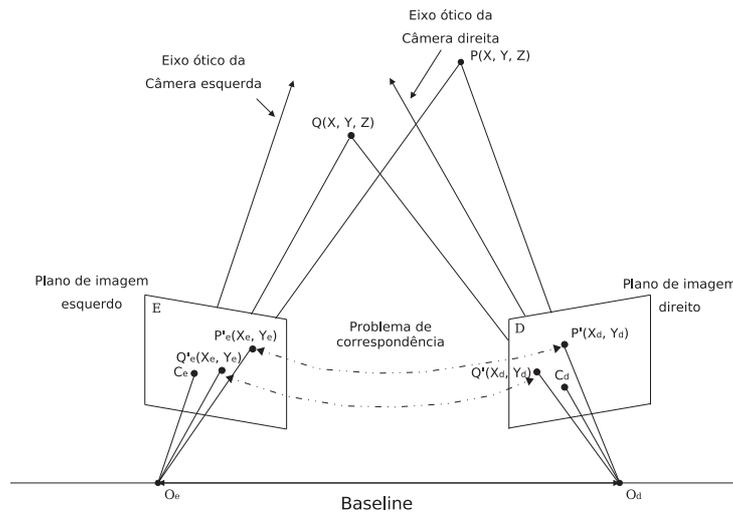
$$Desq = -Ddir \quad (2.3)$$

Nas FIG. 2.4 é apresentado o problema de correspondência a ser resolvido pelos métodos de estimação de disparidades. Na FIG. 2.4a todos os pontos devem ter sua disparidades estimada, mas para ilustrar o problema em questão, dois pontos foram escolhidos (pontos Q e P). Temos na FIG. 2.4b as projeções dos pontos Q e P nos planos de imagem das câmeras esquerdo e direito e a correspondência entre os mesmos estabelecida através de técnicas de estimação de disparidades.

Um mapa de disparidades é uma representação dos vetores de disparidades contendo os valores das diferenças de posições dos pontos homólogos nas vistas direitas e esquerdas do par estéreo. Os valores das disparidades são representados em uma escala de cinza,



(a) Par estéreo 0. Sequência MAN.



(b) Correspondência entre os pontos P e Q.

FIG. 2.4: Problema de correspondência.

onde o nível de intensidade mínimo corresponde ao menor valor de disparidade entre as vistas, e o nível máximo ao maior valor de disparidade.

Com o intuito de desenvolver um método que simule o alto nível de habilidade do SVH na solução do problema de correspondência, muitos métodos de estimação de disparidades foram propostos nas últimas décadas (PEREZ, 2003). Em (BARNARD, 1982; DHOND, 1989; KLETTE, 1998; REDERT, 1999; HARTLEY, 2000; FAUGERAS, 2001; BROW, 2003) encontra-se uma boa revisão sobre imagem estéreo e métodos de estimação de disparidade.

Os principais métodos são baseados em aspectos tais como: área (FAUGERAS, 1993b; KANADE, 1994; ZABIH, 1994; ACCAME, 1995; HENDRIKS, 1996; BHAT, 1998), fluxo óptico (HORN, 1981, 1986), características da imagem (VENKATESWAR, 1995; RAN-

DRIAMASY, 1991; BIRCHFIELD, 1999), fase (MAIMONE, 1996) e otimização da função de energia ou erro (OHTA, 1985; TOMASI, 1995; CORMEN, 1990; KONRAD, 1992).

Os métodos de estimação de disparidades baseados em área são também conhecidos como casamento em bloco ou baseado em bloco.

Para estabelecer o casamento entre os pontos homólogos dos planos de imagem esquerdo e direito, o método usa medidas que se baseiam na similaridade das intensidades dos pixels vizinhos, definidos através de uma janela ao redor do ponto no plano de imagem esquerdo e também ao redor do ponto no plano de imagem direito, cuja correspondência com o primeiro se quer testar. Medidas como SDA (Somatório das Diferenças Absolutas) e SDQ (Somatório das Diferenças ao Quadrado) são utilizadas para quantificar a similaridade da intensidade dos pixels vizinhos nestes métodos (PEREZ, 2003). Tais métodos produzem um denso mapa de disparidades com valores confiáveis nas regiões de textura. Como eles usam correlação entre o somatório dos valores de intensidade dos vizinhos locais do pixel na análise da imagem de referência com o somatório dos valores das intensidades dos pixels vizinhos ao pixel candidato a pixel homólogo na imagem alvo, tais métodos são sensíveis a distorções perspectivas, variações de contraste e iluminação (DHOND, 1989). Entretanto, eles falham nas áreas de oclusão, regiões de baixa textura e não são bons localizadores de descontinuidades (PAGLIARI, 1999).

Algoritmos de disparidades baseados em fluxo óptico, conhecidos também como métodos baseados no gradiente, fazem uso de uma equação diferencial, relacionando o deslocamento e o brilho, para determinar a menor disparidade local entre duas imagens (PEREZ, 2003).

Métodos de estimação de disparidades baseados em características da imagem, convertem as imagens esquerda e direita para uma configuração de características reduzida, como por exemplo, bordas e junções. Essas características na imagem esquerda são casadas com as correspondentes na imagem direita.

Esse tipo de método compõe um mapa de disparidades mais suave do que o algoritmo baseado em área, pois o mapa de disparidades gerado é interpolado para gerar as disparidades dos pixels remanescentes (PEREZ, 2003). Embora provendo uma robusta localização de descontinuidade ou profundidade, esse método é propenso a falha com imagens que não geram uma configuração de características definidas (HOOF, 1989).

Outra aproximação para o cálculo da disparidade binocular são os métodos baseados em fase ou frequência (SANGER, 1988; JENKIN, 1994; MAIMONE, 1995; PAPADIMITRIOU, 1995). Esses métodos empregam a relação entre o deslocamento espacial da

imagem e a mudança de fase da imagem transformada para determinar a correspondência entre as regiões das imagens em cada par (PEREZ, 2003).

Já os métodos baseados em otimização da função de energia ou erro, associada com várias considerações globais das imagens, também conhecidos como métodos de correspondência global, utilizam a informação global da imagem para a estimação. Com isso, apresentam uma complexidade computacional maior do que os outros métodos. Eles empregam algumas técnicas para melhorar o seu desempenho, tais como: programação dinâmica, curvas intrínsecas, graph cuts e métodos bayesianos, sendo o graph cut (CORMEN, 1990), uma das técnicas a serem utilizadas no quarto método desta dissertação (Seção 2.3.2 e Seção 3.5).

### 2.3.1 PROBLEMAS NA ESTIMAÇÃO DE DISPARIDADES

A estimação de disparidades pode ser dificultada por alguns fatores, tais como: o tipo de imagem a ser processada, os ângulos de captura das imagens e a intensidade luminosa. Abaixo são descritos alguns problemas que podem ocorrer na estimação de disparidade (FAUGERAS, 1993a).

- Variação Fotométrica
  - A intensidade de luz emitida ou refletida pelos pontos da cena do mundo real e projetada nos planos de imagem das câmeras tem relação com o posicionamento das câmeras em relação à cena. Desta forma, quando há um deslocamento da câmera do ponto original, ou quando duas câmeras gravam uma cena do mundo real 3D de posições diferentes, as intensidades emitidas ou refletidas pelos pontos da cena do mundo real 3D podem ser diferentes.
  
- Repetição de textura
  - Imagens que apresentam regiões com texturas similares podem gerar no processo de estimação de disparidades múltiplas correspondências. Caso ocorra variação fotométrica este problema é aumentado, podendo levar a uma correspondência errada.

- Falta de textura
  - Nesta seção são apresentadas várias técnicas de estimação de disparidade. Caso a técnica utilizada seja baseada nas características da imagem, nas áreas com ausência de textura (áreas homogêneas da imagem) o estimador poderá executar uma estimação errada ou não encontrar o ponto homólogo.
- Distorção de Perspectiva
  - Este problema é enfatizado quando a *Baseline* possui um valor muito grande. A projeção do objeto da cena do mundo real 3D no plano de imagem esquerdo aparece com o seu formato geométrico diferente comparado ao formato projetado no plano de imagem direito, dificultando a solução do problema de correspondência.
- Oclusão.
  - Pode acontecer de um objeto aparecer em um dos planos de imagem das câmeras e ficar parcialmente ou totalmente ocluído no outro plano de imagem. A oclusão é a dificuldade mais comum que todos os algoritmos de estimação de disparidade enfrentam.  
A FIG. 2.5 mostra a oclusão provocada nas duas vistas pelo objeto colocado no plano de imagem posterior.

### 2.3.2 ESTIMADOR DE DISPARIDADES USANDO O MÉTODO GRAPH CUT

No quarto método de codificação de vídeo estéreo implementado e avaliado nesta tese (Seção 3.5) é utilizado um estimador de disparidades que usa o método graph cut. O estimador de disparidade usando o método graph cut foi escolhido por estar entre os métodos de melhor desempenho propostos na literatura (SCHARSTEIN, 2002). Até a presente data, não há na literatura nenhum codificador de vídeo estéreo que utilize o método graph cut. Esse método é baseado em otimização da função de energia associada com algumas considerações globais da imagem, sendo também conhecido como método de correspondência global (KOLMOGOROV, 2002). Este tipo de método utiliza no processo de estimação informações gerais da imagem. Ao contrário de alguns métodos, este

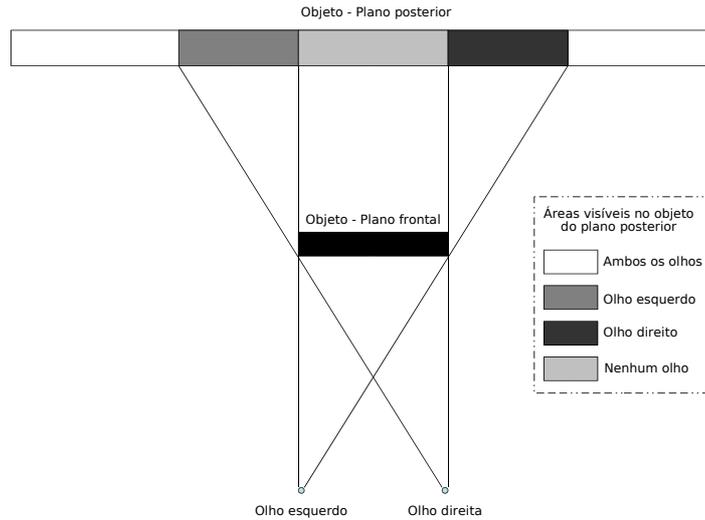


FIG. 2.5: Problema de Oclusão.

analisa toda a imagem em um contexto global, tentando estimar simultaneamente a correspondência de um campo denso de pixels. Isto na teoria aumenta a acurácia (PEREZ, 2003).

Como esse método utiliza considerações globais da imagem, sua complexidade computacional é maior do que a de métodos que utilizam considerações locais para realizar o processo de estimação de disparidade. Por isso, é usado como técnica matemática na localização dos pontos de máximo e mínimo o graph cut, no intuito de diminuir a complexidade computacional. Esse método associa uma função de energia ao mapa de disparidade. Essa função é formada por vários termos citados abaixo que correspondem a uma consideração global (KOLMOGOROV, 2002):

- Fotométrica,
- Suavização e
- Visibilidade.

O método utilizado faz uma aproximação do problema de correspondência através de um ponto de vista de minimização de energia (KOLMOGOROV, 2002).

O graph cut apresenta ótimo desempenho em regiões da imagem que possuem muita textura e tem um desempenho inferior nas regiões da imagem com pouca textura e/ou homogêneas (SCHARSTEIN, 2002).

A energia a ser minimizada tem três propriedades importantes (KOLMOGOROV, 2002):

- Trata as imagens de entrada simetricamente,
- lida com a propriedade de visibilidade, e
- impõe uma suavização espacial enquanto preserva também as descontinuidades.

A FIG. 2.6 mostra a projeção de três pontos de uma cena do mundo real (3D) nos planos de imagens das câmeras esquerda e direita. Onde  $L_n$  (com  $n = 1, 2, 3, 4, 5, 6$ ) correspondem aos planos no espaço 3D afastados das câmeras por uma distância ou profundidade dada por  $n$ . Notamos que os pontos  $P(X, Y, Z)$  e  $S(X, Y, Z)$  pertencem ao mesmo plano no espaço 3D ( $L_3$ ), pois possuem o mesmo valor de profundidade ( $n$ ), sendo os mesmos visíveis as duas câmeras. Vemos que o ponto  $Q(X, Y, Z)$  encontra-se em um plano no espaço 3D ( $L_5$ ) a uma profundidade superior em relação ao ponto  $P(X, Y, Z)$  ( $L_3$ ). Estando o ponto  $Q(X, Y, Z)$  ocluído pelo ponto  $P(X, Y, Z)$  para a câmera direita.

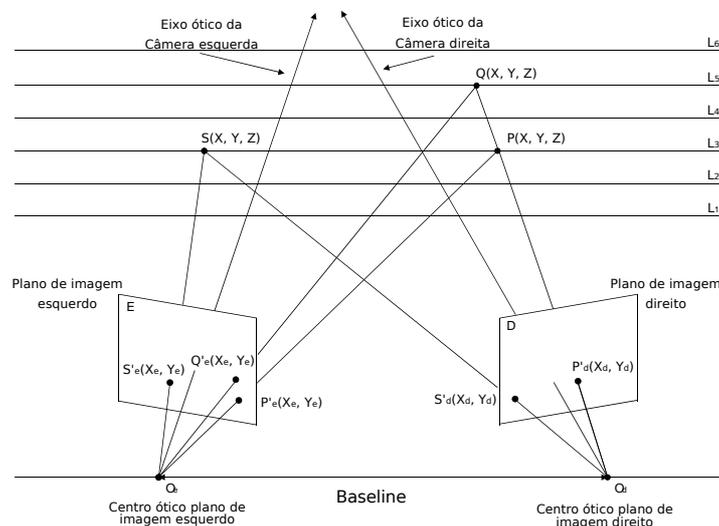


FIG. 2.6: Projeção Graph Cut.

O objetivo do estimador de disparidades é encontrar os planos ( $L_n$ ) no espaço 3D

correspondentes a todos os pontos da cena do mundo real (3D) que se projetam nos planos de imagens das câmeras esquerda e direita.

Sendo um rótulo  $f$  representado unicamente por uma partição de pixels de imagem  $P = Pl \mid l \in L$  onde  $L$  é uma configuração discreta de rótulos ( $l$ ) correspondendo a distâncias ou profundidades (intervalo de valores de disparidade) diferentes e  $Pl = p \in P \mid f_p = l$  é um subconfiguração de pixels assinalados com um rótulo  $l$  (KOLMOGOROV, 2002; BOYKOV, 2001). Um ponto 3D se refere a um par  $(p, l)$  onde  $p \in P, l \in L$  (KOLMOGOROV, 2002).

A função de energia a ser minimizada consiste de três termos (KOLMOGOROV, 2002):

$$E(f) = E_{dados}(f) + E_{suavizacao}(f) + E_{visibilidade}(f) \quad (2.4)$$

O termo dados dará a foto-consistência, conforme abaixo:

$$E_{dados}(f) = \sum_{(p,f(p)),(q,f(q)) \in I} D(p, q) \quad (2.5)$$

onde

$$Dp, q = \min \{0, (Intensity(p) - Intensity(q))^2 - K\} \quad (2.6)$$

$D(p, q)$  é um valor não-positivo dependendo das intensidades dos pixels  $p$  e  $q$  (KOLMOGOROV, 2002), sendo  $K$  uma constante maior que zero.  $I$  é uma configuração de interações (KOLMOGOROV, 2002) consistindo de pares (desordenados) de pontos 3D  $(p_1, l_1)$ ,  $(p_2, l_2)$  próximos um do outro no espaço 3D (pontos  $S$  e  $P$ ), obedecendo a seguinte consideração (KOLMOGOROV, 2002):

- Somente pontos na mesma profundidade podem interagir. Se  $(p_1, l_1)$ ,  $(p_2, l_2) \in I$  então  $l_1 = l_2$ .

O termo de suavização envolve uma noção de vizinhança, em (KOLMOGOROV, 2002) é assumido que exista um sistema de vizinhança sobre os pixels

$$N \subset \{\{p, q\} \mid p, q \in P\} \quad (2.7)$$

quando, para um ponto  $p = (p_x, p_y)$  e  $q = (q_x, q_y)$  obedecem a igualdade

$$|p_x - q_x| + |p_y - q_y| = 1 \quad (2.8)$$

sendo o termo de suavização reescrito como

$$E_{suavizacao}(f) = \sum_{(p,q) \in N} V_{p,q}(f(p), f(q)) \quad (2.9)$$

O último termo refere-se à consideração de visibilidade. Caso a consideração seja satisfeita, este termo terá valor zero, caso contrário será igual a  $\infty$ .

Pares de pontos 3D que violam a consideração de visibilidade pertencem a uma configuração de interações  $I_{vis}$ , conforme a condição abaixo:

- Somente pontos 3D em profundidades diferentes podem interagir. Se  $(p_1, l_1)$ ,  $(p_2, l_2) \in I$  então  $l_1 \neq l_2$ .

$$E_{visibilidade}(f) = \sum_{(p,f(p)),(q,f(q)) \in I_{vis}} \infty \quad (2.10)$$

Pesquisas recentes mostram que algoritmos rápidos de minimização de energia associadas a características da imagem têm sido desenvolvidos baseados em graph cuts (BOYKOV, 2001; KOLMOGOROV, 2001; ROY, 1999; BOYKOV, 1998; ISHIKAWA, 1998; ROY, 1998).

Um melhor detalhamento do método é descrito em (KOLMOGOROV, 2002).

## 2.4 MONITORES ESTEREOSCÓPICOS.

A exibição de imagens ou vídeos estereoscópicos necessita de monitores especiais (WOJCIECH, 2004; DAVI, 2003; LIPTON, 1991). Segundo (WOODS, 2004; DAVI, 2003), durante os últimos anos, várias empresas pesquisaram e desenvolveram tecnologias para a fabricação de monitores estéreo. Várias patentes são concedidas a cada ano, mas nem todas as propostas são comercialmente viáveis. A tecnologia de monitores estereoscópicos tem feito grandes avanços em poucos anos. Além da melhora na tecnologia utilizada nos monitores, eles estão consideravelmente baratos.

Pesquisas realizadas por Tam e Stelmach (TOMASI, 1995) mostram que 96% dos jovens telespectadores podem perceber a sensação de estéreo. Desta forma, se não há problema e desconforto para os mesmos, a utilização de monitores estereoscópicos é viável.

Pesquisadores também têm se beneficiado dessa nova tecnologia, sendo a mesma empregada em estudos nas áreas de modelagem molecular, simulações de vôo, CAD, medida de áreas e edição de mapas e gráficos através de fotografias aéreas (fotogrametria), visualização de dados multidimensionais, imagens médicas, tele-operações, estereolitografia e arte (DAVI, 2003; WOJCIECH, 2004).

Segundo (WOJCIECH, 2004), sistemas de TV 3D são esperados como a próxima revolução na história da televisão. E quando falamos de vídeo 3D, geralmente nos referimos ao armazenamento de seqüências animadas, mas quando nos referimos a TV 3D, isto inclui aquisição em tempo real, codificação, e transmissão de cenas dinâmicas. Em (WOJCIECH, 2004) é apresentado o primeiro sistema de TV 3D fim-a-fim com 16 câmeras independentes, vistas de alta-resolução e monitores autoestereoscópicos.

Existem vários tipos de monitores 3D para a exibição da imagem ou vídeo estéreo. Referências adicionais podem ser encontradas em (WOODS, 2004; WOJCIECH, 2004; DAVI, 2003).

## 2.5 RESUMO

Neste capítulo, foram apresentados os conceitos básicos de estereoscopia abordados ou empregados nesta dissertação. Neste contexto foram apresentados os conceitos de disparidade estéreo, métodos de estimação de disparidades e o estimador de disparidades usando o método graph cut utilizado no quarto método avaliado (Seção 3.5) juntamente com comentários sobre monitores estéreo.

### 3 CODIFICAÇÃO DE VÍDEO ESTÉREO

Neste capítulo são apresentados conceitos básicos de codificação de imagem/vídeo, características do codificador utilizado nas simulações e os quatro métodos de codificação de vídeo estéreo avaliados. Na Seção 3.1 é feita uma breve introdução sobre vídeo estéreo. Conceitos relevantes de codificação de uma imagem e vídeo monocular são apresentados na Seção 3.2 e Seção 3.3. Na Seção 3.4 é apresentado o codificador H.264/AVC, sua estrutura e suas principais características. Os quatro métodos de codificação de vídeo estéreo estudados são apresentados na Seção 3.5. Por fim na Seção 3.6, apresenta-se o resumo do capítulo.

#### 3.1 INTRODUÇÃO

Como o vídeo estéreo é composto de mínimo duas seqüências de quadros formadas pela captura de uma cena do mundo real 3D através de ângulos diferentes, a princípio o mesmo gera o dobro da taxa de bits comparado ao sistema monocular, quando codificado como duas seqüências monoculares. Há a necessidade de se utilizar um codificador que obtenha um alto desempenho no processo de codificação das seqüências estéreo, para que as mesmas possam ser utilizadas em um grande número de aplicações (PAGLIARI, 1999). Vários métodos já foram propostos com o intuito de obter uma codificação eficiente para imagens e seqüências estéreo (DUARTE, 2005; YAN, 2003; ELLINAS, 2005; SUN, 2003; LIN, 1999; PEI, 1998; TZOVARAS, 1997; PERKINS, 1992). Nesta dissertação são estudados, avaliados e comparados, quatro métodos de codificação de vídeo estéreo (Seção 1.2). O quarto método avaliado faz uso de um estimador de disparidades com o objetivo de tentar diminuir a taxa de bits resultante da codificação da seqüência estéreo.

#### 3.2 CODIFICAÇÃO DE IMAGEM

Na FIG. 3.1, temos o diagrama de blocos de um codificador genérico com perdas.

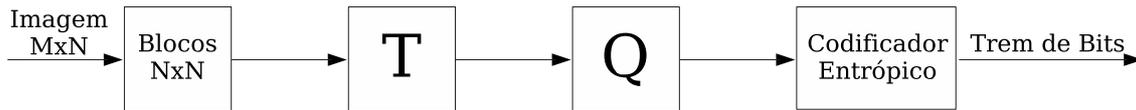


FIG. 3.1: Estrutura de um codificador genérico.

A finalidade de um codificador com ou sem perdas é explorar as redundâncias espaciais da imagem e as redundâncias do código entrópico (*Huffman* ou codificação aritmética). No codificador com perdas, a imagem reconstruída nunca é igual à original. Em geral, num processo de codificação com perdas a imagem é decomposta ou transformada, quantizada e codificada. Quando o codificador utiliza a transformada cosseno discreta (DCT) como método de decorrelação dos dados da imagem, esta é dividida em blocos  $N \times N$  e cada bloco  $N \times N$  tem seus pixels transformados (T) do domínio da imagem para o domínio da transformada (LIM, 1990). A DCT é a transformada baseada em blocos mais utilizada na codificação de imagens e vídeo (RICHARDSON, 2003).

A DCT inversa tem a mesma implementação da direta, pois a transformada cosseno é real e ortogonal (JAIN, 1989; LIM, 1990). A DCT possui a propriedade de compactação de energia. Para uma imagem altamente correlatada, a DCT concentra em alguns poucos coeficientes (DC e AC de mais baixa ordem) a energia da imagem, onde o coeficiente DC representa o nível médio dos dados 2D (JAIN, 1989).

A formação de bordas horizontais, verticais e diagonais na imagem resulta da variação brusca da intensidade dos pixels vizinhos, nestas direções. A FIG. 3.2 ilustra a distribuição dos coeficientes transformados de um bloco  $N \times N$ , de acordo com a variação da intensidade dos pixels.

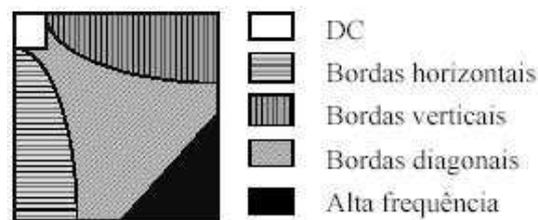


FIG. 3.2: DC, Bordas e Altas Frequências - DCT. (PAGLIARI, 1999).

É importante ressaltar que os blocos podem apresentar classes de atividades diferentes, i.e., há blocos com pouca variação de intensidade e outros com muita variação. A DCT consegue compactar em poucos coeficientes blocos com baixa variação (dados correlatados). No entanto, blocos com alta variação de intensidade dos pixels vizinhos apresentam uma distribuição de coeficientes no domínio da DCT, de acordo com a intensidade e direção dessa variação (FIG. 3.2).

Na transformação do bloco  $N \times N$  pela DCT não há compressão na informação e sim uma compactação da energia do bloco  $N \times N$  em poucos coeficientes de mais baixa ordem e com muitos coeficientes contendo pouca energia no domínio da DCT. O bloco  $N \times N$  de coeficientes transformados é enviado ao quantizador (Q) onde será realizada a alocação de bits. De maneira geral, como o SVH é menos sensível a distorções nas áreas de alta frequência na imagem, mais bits podem ser alocados para os coeficientes de mais baixa ordem (baixa frequência), enquanto para os que representam os coeficientes de alta frequência no domínio da DCT pode-se alocar poucos ou nenhum bit, sem grandes prejuízos à imagem reconstruída (JAIN, 1989). No entanto, o objetivo de um sistema de compressão com perdas é a obtenção de distorção mínima para uma taxa de bits (controle de taxa-distorção). Além dos métodos de codificação por transformadas (e.g. DCT), os métodos de codificação por sub-bandas, especialmente usando wavelets (DEROSE, 1994; HILTON, 1994), têm-se mostrado bastante eficientes na compressão de imagens/vídeos (M. D. ADAMS, 2000; ISO, 2002; ANTONINI, 1992). Apesar da teoria taxa-distorção não levar em conta o SVH a fim de prover a qualidade (distorção) desejada para uma determinada taxa de bits, pode-se utilizar um processo de alocação de bits que leve em conta o SVH (alocação de bits perceptual). Neste caso, os degraus de quantização podem ser determinados de acordo com a sensibilidade do SVH (SULLIVAN, 2004).

Quanto maior for a quantidade de bits alocada a um coeficiente, maior será a quantidade de níveis no quantizador para a representação do mesmo. O bloco de coeficientes quantizado é então enviado ao último estágio, que é a codificação por entropia, onde o trem de bits é formado tendo no seu cabeçalho informações importantes para o decodificador, como por exemplo: as tabelas de quantização utilizadas (tabela de luminância e crominância) e a tabela gerada pelo codificador de entropia.

Para ilustrar a propriedade de compactação de energia da DCT, A imagem de teste Lena (FIG. 3.3a) foi dividida em blocos  $32 \times 32$ , onde cada bloco foi transformado pela DCT. A FIG. 3.3a é formada por regiões homogêneas e de baixa, média e alta frequência.

As regiões homogêneas e de baixa frequência são encontradas no rosto, no ombro e em algumas partes do fundo da figura, as de média frequência estão localizadas no chapéu e na sua aba, e as de alta frequência estão localizadas nas plumas. A FIG. 3.3b ilustra os blocos transformados pela DCT. Os pontos brancos representam os coeficientes DC dos blocos e os pontos em tons de cinza, os coeficientes AC da DCT do bloco. Esta representação mostra os coeficientes de maior energia em tons mais claros e os de menor energia em tons mais escuros (o ponto de maior energia representado em branco e o de menor energia em preto). Dois blocos foram escolhidos para análise de seus coeficientes, sendo um deles com muitos detalhes (bloco 1) e o outro com poucos detalhes (bloco 2). Podemos perceber que no bloco 1 (FIG. 3.3c), em que há muitas variações bruscas da intensidade, não podemos desconsiderar (quantizar grosseiramente) muitos coeficientes, pois perderíamos informação. Porém, no bloco 2 (FIG. 3.3d), há muitas regiões homogêneas, o que, como visto anteriormente, possibilita um bom nível de compressão da imagem.

Os coeficientes dos blocos 1 e 2 são exibidos em representação tridimensional, FIG. 3.3e e FIG. 3.3f. Nestas representações, os valores dos coeficientes DC (coordenada 0,0) foram manipulados (escala diferente) para que os coeficientes AC fossem mais bem visualizados. O objetivo destas figuras é mostrar que regiões de pouca textura são "facilmente" comprimidas, enquanto regiões contendo muita textura são mais "difíceis" de serem comprimidas. Este conceito será importante quando for apresentado o quarto método de codificação estéreo que usa um estimador de disparidades (Seção 3.5).

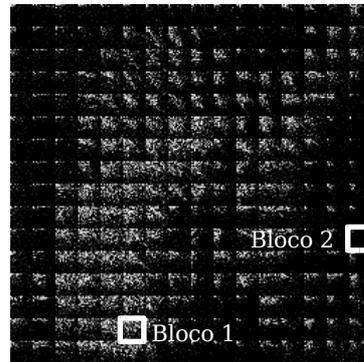
### 3.3 CODIFICAÇÃO DE VÍDEO MONOCULAR

Seqüências monoculares são largamente utilizadas atualmente, sendo perfeitamente aceitáveis quando a informação de profundidade não é necessária. Codificadores como MPEG-1, MPEG-2, H.263 e H.264/AVC (ISO, 1993; ITU-T, 1994, 1995, 2005) foram projetados para codificar seqüências monoculares. Na FIG. 3.4, temos a formação de uma seqüência monocular.

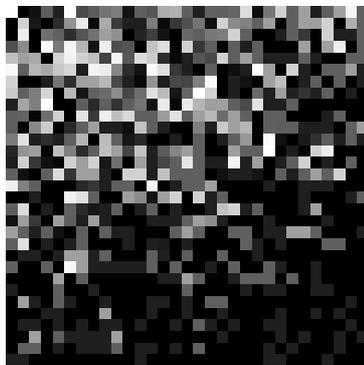
O processo de codificação de vídeo é geralmente constituído de uma codificação de vídeo híbrida baseada em blocos. Sendo o algoritmo de codificação do H.264/AVC um híbrido de predição inter-quadro para explorar as redundâncias temporais e uma codificação por transformada para explorar as redundâncias espaciais do resíduo (imagem de erro) predito.



(a) Imagem Original



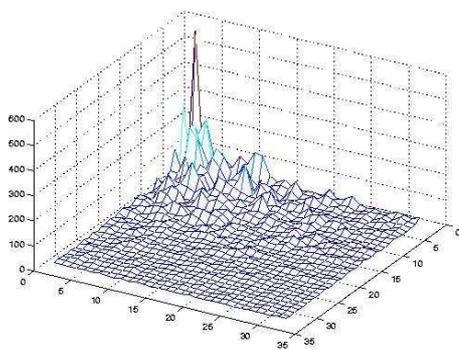
(b) DCT com Blocos 32x32



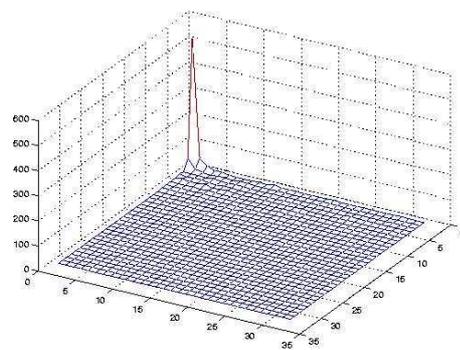
(c) DCT do Bloco 1



(d) DCT do Bloco 2



(e) Bloco 1 - Representação 3D



(f) Bloco 2 - Representação 3D

FIG. 3.3: Imagem de teste Lena 512x512



FIG. 3.4: Formação de uma Seqüência Monocular

Na primeira etapa do processo de codificação o quadro é dividido em pequenas áreas (FIG. 3.5) denominadas de macroblocos. Cada macrobloco é constituído por blocos de 16x16 amostras de luminância (Y) associadas com blocos de 8x8 amostras de crominância das componentes azuis (Cb) e vermelhas (Cr) de acordo com a sub-amostragem utilizada (SULLIVAN, 2004).

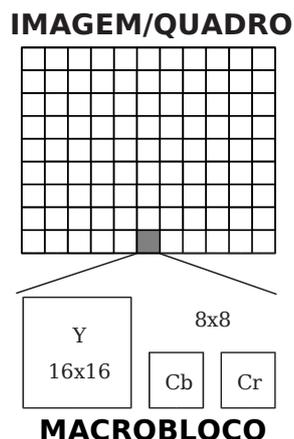


FIG. 3.5: Quadro dividido em macroblocos. (SCHAFER, 2003; WIEGAND, 2003).

Quando o codificador explora a redundância espacial de um quadro é realizada uma predição intra-quadro. Um quadro formado somente com macroblocos de uma predição intra-quadro recebe o nome de quadro do tipo I. Seqüências monoculares apresentam, na maioria das vezes, uma correlação muito grande entre o quadro anterior e o quadro atual, possibilitando a aplicação de uma técnica que se vale da exploração da redundância temporal (inter-quadros) para diminuir a taxa de bits da seqüência codificada. Esta técnica

é chamada de estimação de movimento. Este processo de estimação de movimento é baseado em blocos. Na FIG. 3.6, o processo de estimação de movimento de um macrobloco (quadrado branco) é ilustrado (FIG. 3.6b). O estimador procura em uma área dentro da imagem de referência (quadros anteriores ou posteriores) para encontrar a região de melhor casamento. Essa região é comparada com o macrobloco da imagem atual. Algumas ou todas as regiões dentro de uma área de busca, de tamanho pré-definido (usualmente a região é centrada sobre a posição do macrobloco corrente, FIG. 3.6a) são comparadas, e a região escolhida será aquela com o melhor casamento. Um critério bastante utilizado de casamento é a energia da diferença entre uma região dentro da área de busca e o macrobloco da imagem corrente (resíduo da estimação de movimento). Desta forma, a região que minimizar a energia do resíduo, será a escolhida. Para cada macrobloco no quadro de referência é eleito o seu melhor correspondente no quadro atual. A região escolhida é predita do macrobloco da imagem corrente (RICHARDSON, 2003). Para cada par de macroblocos nos quadros de referência e atual é gerado o resíduo resultante da estimação de movimento. O processo de estimação de movimento e predição se repete para todos os macroblocos do quadro de referência gerando uma imagem de resíduos da estimação de movimento, conhecida como imagem de erro (FIG. 3.6c). Este processo ocorre somente quando o codificador decide utilizar a estimação de movimento em vez da predição intra-quadro.

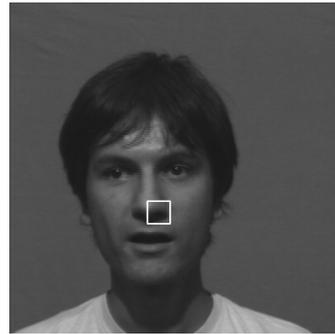
Se a busca do macrobloco é realizada tendo como quadros de referência os que tenham sido codificados anteriormente ao quadro atual, o quadro é denominado de tipo P. Caso a busca seja feita usando como referência os quadros anteriores e posteriores ao quadro atual, o mesmo é denominado de tipo B.

A predição a ser executada no quadro atual obedece a uma configuração pré-estabelecida no codificador denominada de GoP (Group of Pictures). Na FIG. 3.7, temos um exemplo de GoP, onde o valor de (N) define o tamanho do GoP (intervalo entre quadros tipo I) e o valor (M) define o intervalo de quadros tipo P.

Quanto mais preciso for o processo de estimação de movimento mais homogêneo será o resíduo. Desta forma, maior será a compressão feita no resíduo pois o mesmo, após a transformada, terá poucos coeficientes com informação relevante. Assim, o quantizador poderá alocar bits só para os coeficientes com informação relevante e desconsiderar (ou alocar poucos bits) para os demais. Se o macrobloco for totalmente homogêneo, o quantizador só precisará alocar bits para o coeficiente DC, zerando os demais sem perda de



(a) Quadro 4 - Referência



(b) Quadro 5 - Atual



(c) Imagem de erro - Quadros 4-5

FIG. 3.6: Seqüência MAN - Estimaco de Movimento.

informaco para a reconstruco do macrobloco do decodificador.

Entretanto, gerar um resduo o mais homogneo possvel no garante uma taxa de codificaco inferior, pois, dependendo do tipo de seqüncia a ser codificada, pode haver a necessidade de utilizar um nmero elevado de vetores de movimento. As informaces referentes aos mesmos so transmitidas juntamente com os dados residuais codificados no trem de bits, gerando um aumento no mesmo. Ou seja, nem sempre o melhor casamento de macroblocos produz o melhor compromisso taxa-distorco, uma vez que para atingir uma correspondncia tima h o custo de se utilizar um elevado nmero de vetores de movimento. J no quarto mtodo de codificaco de vdeo estreo, que utiliza um estimador de disparidades, a obteno de um resduo mais homogneo est diretamente associada a um melhor desempenho na taxa de codificaco, juntamente com a qualidade da seqüncia reconstruda, como veremos na Seo 3.5.

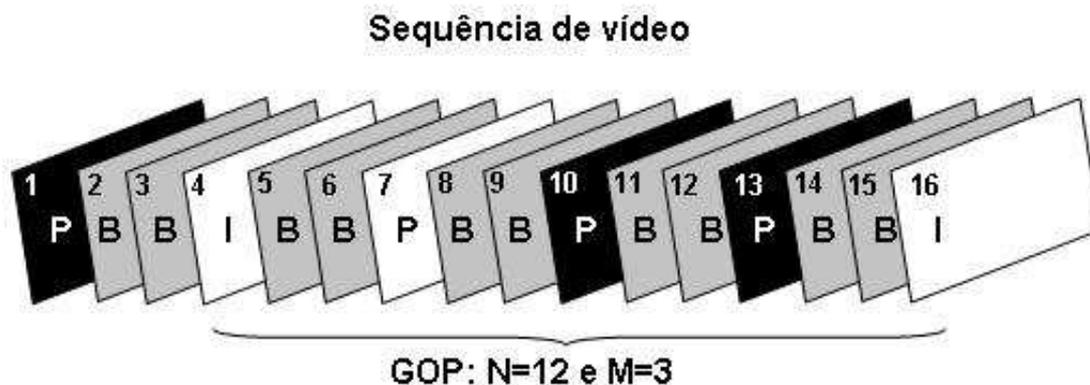


FIG. 3.7: GOP com quadros I, P e B.  $M=3$  e  $N=12$ .

### 3.4 O CODIFICADOR H.264/AVC

O codificador H.264/AVC (Advanced Video Coding) é o mais recente padrão internacional de codificação de vídeo desenvolvido pelo Moving Pictures Expert Group (MPEG) e pelo Video Coding Expert Group (VCEG).

Em 2001, o VCEG uniu-se ao MPEG formando o Joint Video Team (JVT), no intuito de finalizar o proposto H.26L (esboço de um novo padrão para comunicações visuais com taxa de bits baixa) e converter o mesmo para um padrão internacional de codificação de vídeo. Em março de 2005, o novo padrão internacional de codificação de vídeo foi aprovado, gerando três recomendações: H.264/AVC (Advanced Video Coding for generic audiovisual services), H.264.1 (Conformance specification for H.264/AVC advanced video coding) e H.264.2 (Reference software for H.264/AVC advanced video coding) (ITU-T, 2005). Por ele representar o estado da arte entre os demais codificadores presentes atualmente no mercado (SCHAFER, 2003; WIEGAND, 2003; HONG, 2002; RICHARDSON, 2002) foi escolhido para ser utilizado nesta dissertação na codificação de seqüências estéreo. Ele tem como objetivos principais melhorar o processo de compressão e prover uma interface amigável para uma variedade de camadas de transporte e armazenamento de mídias. Aplicações em várias áreas têm se beneficiado do novo padrão, tais como (RICHARDSON, 2003; WIEGAND, 2003):

- Serviços de mensagens multimídia sobre ISDN, DSL, Ethernet, LAN sem fio e sistemas móveis.
- Vídeo sobre demanda ou serviços de multimídia sobre trem de bits em ISDN, modem por cabo, DSL, LAN, sistemas sem fio.
- Serviços conversacionais sobre ISDN, Ethernet, LAN, DSL, sistemas móveis e sem fio, modem, etc, ou uma misturas desses.
- Armazenamento serial ou interativo sobre equipamentos ópticos ou magnéticos, DVD.
- Broadcast sobre cabo, satélite, modem por cabo, DSL, terrestre.

### 3.4.1 ESTRUTURA DO CODIFICADOR H.264/AVC

A estrutura do codificador H.264/AVC é composta por duas camadas distintas. A primeira camada é denominada de camada de codificação de vídeo (CCV) (RICHARDSON, 2003; WIEGAND, 2003) sendo a mesma responsável pelo processo de codificação dos quadros da seqüência de vídeo que chegam à entrada do codificador. Os processos de predição, transformada, quantização e codificação por entropia dos quadros da seqüência de vídeo compõem esta camada. Os quadros codificados da seqüência são enviados em um trem de bits para a próxima camada do codificador, denominada camada de abstração de sistema (CAS). Nesta camada, o trem de bits que representa o conteúdo de vídeo da seqüência codificada é provido de um cabeçalho apropriado para uma camada de transporte, no caso de uma transmissão, ou de um determinado tipo de mídia para armazenamento.

### 3.4.2 CAMADA DE CODIFICAÇÃO DE VÍDEO (CCV)

O alto ganho de compressão obtido pelo codificador não resulta de uma melhoria em um determinado estágio do processo de codificação e sim de pequenas melhorias em todos os estágios da camada de codificação de vídeo, pois o mesmo apresenta os mesmos elementos funcionais dos padrões de codificação de vídeo anteriores.

Na FIG. 3.8, temos o diagrama em blocos de um codificador H.264/AVC.

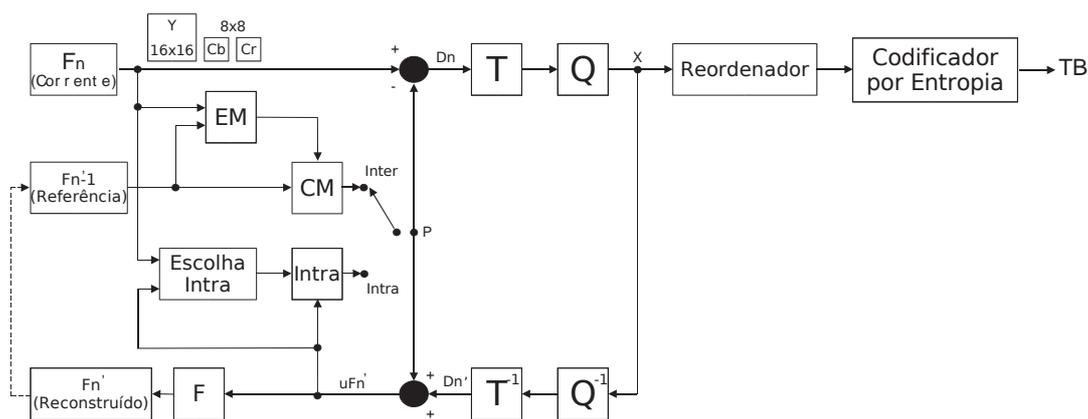


FIG. 3.8: Diagrama do codificador H.264/AVC (SCHAFER, 2003; WIEGAND, 2003)

No primeiro estágio do codificador (FIG. 3.8) o quadro é dividido em macroblocos para que, através de técnicas de predição, suas redundâncias espaciais e temporais sejam exploradas. Os resíduos obtidos no processo de predição intra-quadro ou inter-quadro são enviados ao segundo estágio, onde são transformados ( $T$ ) e quantizados ( $Q$ ). Ainda neste estágio é realizada a quantização inversa ( $Q^{-1}$ ) e a transformada inversa ( $T^{-1}$ ) para que os resíduos codificados possam ser utilizados na predição dos próximos quadros. Antes de serem utilizados no processo de predição dos próximos quadros, estes resíduos passam por um filtro de redução de blocagem ( $F$ ). A utilização deste filtro melhora o quadro reconstruído e a acurácia do processo de predição. No último estágio, os macroblocos transformados e quantizados são reordenados e enviados a um codificador por entropia, que gera um trem de bits da seqüência codificada. Este trem de bits é enviado para a camada CAS.

Como foi dito anteriormente, o ganho no processo de codificação do codificador H.264/AVC não é devido a um único elemento funcional, e sim as melhorias obtidas em todos eles.

Uma configuração das ferramentas utilizadas no processo de codificação é denominada perfil. Os fatores que limitam o desempenho para os codificadores e decodificadores, tais como tamanho da imagem, taxa de bits codificada e memória requerida, são denominados de nível (RICHARDSON, 2003). Inicialmente, foram definidos três perfis para o H.264/AVC (*Baseline*, *Main* e *Extend*). Em 2004, o trabalho de pesquisa sobre novas

ferramentas de codificação para estender os três perfis iniciais foi completado, sendo essa extensão denominada de FRExt (*Fidelity Range Extensions*). Esta extensão definiu mais quatro perfis (*High*(HP), *High* 10(Hi10P), *High* 4:2:2(Hi422P) e *High* 4:4:4(Hi444P)). Estes quatro novos perfis foram desenvolvidos baseados no projeto do perfil *Main* (SULLIVAN, 2004) e, no início de 2005, essa extensão foi incorporada ao padrão do H.264/AVC. O padrão H.264/AVC possui dezesseis níveis, correspondendo cada um deste níveis a um limite em relação à taxa de processamento das amostras, ao tamanho da imagem, etc, no processo de codificação e decodificação. Em (RICHARDSON, 2003; SULLIVAN, 2004) encontra-se a relação entre os perfis e as ferramentas de codificação presentes em cada um deles e dos seus níveis e limites.

Os quadros de uma seqüência de vídeo entrelaçada podem ser tratados como quadros progressivos ou quadros formados por dois campos no processo de codificação. O codificador H.264/AVC apresenta três maneiras distintas em relação à forma de codificação dos quadros, da seqüência de vídeo entrelaçado em sua entrada. Ele pode codificar todos os quadros da seqüência de vídeo entrelaçado como quadros progressivos ou quadros formados por dois campos (desmontando o quadro entrelaçado em dois campos) e também de uma forma adaptativa, decidindo através da relação taxa-distorção, a melhor forma de codificação de cada quadro da seqüência, entre progressiva e por campo. Esta última forma de codificação da seqüência entrelaçada é proporcionada pela ferramenta denominada de *Picture Adaptive Frame-Field* (PAFF) (WANG, 2002b). No arquivo de configuração, a *flag PicInterlace* informar ao codificador, qual entre as três maneiras de codificações da seqüência entrelaçada será utilizada. Quando o quadro entrelaçado é codificado por campo, os dois campo são codificados seqüencialmente. As técnicas de predição intra-quadro e inter-quadro utilizadas são as mesmas da codificação por quadro progressivo.

Nesta dissertação, duas ferramentas desenvolvidas para a codificação de seqüências de vídeo entrelaçado, presentes tanto no perfil *Main* e *High*, são analisadas. Os detalhes do processo de predição no modo Intra e Inter destas ferramentas não se encontram na recomendação do codificador H.264/AVC (ITU-T, 2005) e nem em livros sobre este codificador. Com isso, os detalhes destas ferramentas foram obtidos do estudo de dois documentos ((WANG, 2002b,a) PAFF e MBAFF) enviados ao JVT como propostas para essas ferramentas.

A primeira ferramenta é PAFF, que como visto no parágrafo acima, capacita o codificador a decidir se a imagem será codificada como um quadro progressivo ou entrelaçado,

ou seja, se os dois campos do quadro entrelaçado serão combinados e codificados como um quadro progressivo, ou se os mesmos não serão combinados e codificados separadamente como dois campos. Quando o codificador H.264/AVC codifica todos os quadros de uma seqüência de vídeo entrelaçado como progressivo ou entrelaçado, ele não leva em consideração o conteúdo da imagem. Segundo (WIEGAND, 2003), quadros entrelaçados com regiões de movimentos de objetos ou movimentos de câmera tendem a ter suas linhas adjacentes com um grau de dependência estatística reduzido, quando comparado com quadros progressivos. Neste caso, quando se faz uso da codificação adaptativa entre quadro progressivo e entrelaçado, pode-se codificar o quadro entrelaçado como progressivo obtendo uma codificação mais eficiente de seqüências com as características anteriores.

A segunda ferramenta analisada trabalha em conjunto com a PAFF, e é denominada de *Picture MacroBlock Adaptive Frame/Field*(MBAFF) (WANG, 2002a). Esta ferramenta capacita o codificador de realizar a decisão descrita acima, á nível de macrobloco entre os modos de codificação do vídeo entrelaçada progressivo e por campo. Em casos onde a seqüência consiste de quadros possuindo algumas regiões com movimento e outras não, é mais eficiente codificar as regiões sem movimento no modo progressivo e as com movimento no modo entrelaçado.

Na codificação progressiva do quadro entrelaçado, o macrobloco no processo de predição pode ser dividido em sete modos (FIG. 3.9). Os tamanhos dos blocos podem ser 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, ou 4x4. Sendo que na predição intra-quadros, o codificador pode utilizar somente os modos um e sete (FIG. 3.9).

Na codificação por campo do quadro entrelaçado, o macrobloco no processo de predição pode ser dividido em cinco modos (FIG. 3.10). Os tamanhos dos blocos podem ser 8x16, 8x8, 8x4, 4x8, ou 4x4. Sendo que na predição intra-quadros, o codificador pode utilizar somente os modos um e cinco (FIG. 3.10).

Nas codificações progressiva ou por campo, na predição inter-quadros, o codificador pode utilizar qualquer um dos modos (WANG, 2002a).

Podemos observar que o codificador quando opta pela codificação por campo, possui um número de menor de modos para a predição dos quadros. Além disso, este tipo de codificação não disponibiliza de dois modos de tamanho de blocos (16x16 e 16x8) muito populares para muitas seqüências. Desta forma, para que a codificação por campo tenha um bom desempenho em relação à codificação progressiva, a mesmo deveria possuir o mesmo número de tamanhos de blocos no processo de predição dos quadros. A solução

Modo 1	Modo 2	Modo 3	Modo 4								
0	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 50%; height: 50%;">0</td><td style="width: 50%; height: 50%;">1</td></tr> </table>	0	1	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 50%; height: 50%;">0</td></tr> <tr><td style="width: 50%; height: 50%;">1</td></tr> </table>	0	1	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 50%; height: 50%;">0</td><td style="width: 50%; height: 50%;">1</td></tr> <tr><td style="width: 50%; height: 50%;">2</td><td style="width: 50%; height: 50%;">3</td></tr> </table>	0	1	2	3
0	1										
0											
1											
0	1										
2	3										

Modo 5	Modo 6	Modo 7																																
<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 25%; height: 25%;">0</td><td style="width: 25%; height: 25%;">1</td><td style="width: 25%; height: 25%;">2</td><td style="width: 25%; height: 25%;">3</td></tr> <tr><td style="width: 25%; height: 25%;">4</td><td style="width: 25%; height: 25%;">5</td><td style="width: 25%; height: 25%;">6</td><td style="width: 25%; height: 25%;">7</td></tr> </table>	0	1	2	3	4	5	6	7	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 50%; height: 25%;">0</td><td style="width: 50%; height: 25%;">1</td></tr> <tr><td style="width: 50%; height: 25%;">2</td><td style="width: 50%; height: 25%;">3</td></tr> <tr><td style="width: 50%; height: 25%;">4</td><td style="width: 50%; height: 25%;">5</td></tr> <tr><td style="width: 50%; height: 25%;">6</td><td style="width: 50%; height: 25%;">7</td></tr> </table>	0	1	2	3	4	5	6	7	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr><td style="width: 25%; height: 25%;">0</td><td style="width: 25%; height: 25%;">1</td><td style="width: 25%; height: 25%;">2</td><td style="width: 25%; height: 25%;">3</td></tr> <tr><td style="width: 25%; height: 25%;">4</td><td style="width: 25%; height: 25%;">5</td><td style="width: 25%; height: 25%;">6</td><td style="width: 25%; height: 25%;">7</td></tr> <tr><td style="width: 25%; height: 25%;">8</td><td style="width: 25%; height: 25%;">9</td><td style="width: 25%; height: 25%;">10</td><td style="width: 25%; height: 25%;">11</td></tr> <tr><td style="width: 25%; height: 25%;">12</td><td style="width: 25%; height: 25%;">13</td><td style="width: 25%; height: 25%;">14</td><td style="width: 25%; height: 25%;">15</td></tr> </table>	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	2	3																															
4	5	6	7																															
0	1																																	
2	3																																	
4	5																																	
6	7																																	
0	1	2	3																															
4	5	6	7																															
8	9	10	11																															
12	13	14	15																															

FIG. 3.9: Codificação Progressiva - Modos de predição (WANG, 2002a).

Modo 1	Modo 2	Modo 3	Modo 4	Modo 5																																						
<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr style="background-color: #cccccc;"><td style="width: 50%; height: 50%;">0</td></tr> <tr><td style="width: 50%; height: 50%;">1</td></tr> </table>	0	1	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr style="background-color: #cccccc;"><td style="width: 50%; height: 50%;">0</td><td style="width: 50%; height: 50%;">1</td></tr> <tr><td style="width: 50%; height: 50%;">2</td><td style="width: 50%; height: 50%;">3</td></tr> </table>	0	1	2	3	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr style="background-color: #cccccc;"><td style="width: 25%; height: 25%;">0</td><td style="width: 25%; height: 25%;">1</td><td style="width: 25%; height: 25%;">2</td><td style="width: 25%; height: 25%;">3</td></tr> <tr><td style="width: 25%; height: 25%;">4</td><td style="width: 25%; height: 25%;">5</td><td style="width: 25%; height: 25%;">6</td><td style="width: 25%; height: 25%;">7</td></tr> </table>	0	1	2	3	4	5	6	7	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr style="background-color: #cccccc;"><td style="width: 50%; height: 25%;">0</td><td style="width: 50%; height: 25%;">1</td></tr> <tr style="background-color: #cccccc;"><td style="width: 50%; height: 25%;">2</td><td style="width: 50%; height: 25%;">3</td></tr> <tr><td style="width: 50%; height: 25%;">4</td><td style="width: 50%; height: 25%;">5</td></tr> <tr><td style="width: 50%; height: 25%;">6</td><td style="width: 50%; height: 25%;">7</td></tr> </table>	0	1	2	3	4	5	6	7	<table border="1" style="border-collapse: collapse; width: 100%; height: 100%;"> <tr style="background-color: #cccccc;"><td style="width: 25%; height: 25%;">0</td><td style="width: 25%; height: 25%;">1</td><td style="width: 25%; height: 25%;">2</td><td style="width: 25%; height: 25%;">3</td></tr> <tr style="background-color: #cccccc;"><td style="width: 25%; height: 25%;">4</td><td style="width: 25%; height: 25%;">5</td><td style="width: 25%; height: 25%;">6</td><td style="width: 25%; height: 25%;">7</td></tr> <tr><td style="width: 25%; height: 25%;">8</td><td style="width: 25%; height: 25%;">9</td><td style="width: 25%; height: 25%;">10</td><td style="width: 25%; height: 25%;">11</td></tr> <tr><td style="width: 25%; height: 25%;">12</td><td style="width: 25%; height: 25%;">13</td><td style="width: 25%; height: 25%;">14</td><td style="width: 25%; height: 25%;">15</td></tr> </table>	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0																																										
1																																										
0	1																																									
2	3																																									
0	1	2	3																																							
4	5	6	7																																							
0	1																																									
2	3																																									
4	5																																									
6	7																																									
0	1	2	3																																							
4	5	6	7																																							
8	9	10	11																																							
12	13	14	15																																							

FIG. 3.10: Codificação por Campo - Modos de predição (WANG, 2002a).

encontrada foi a introdução do conceito de super macroblocos. Um super macrobloco 32x16 é constituído de dois macroblocos 16x16. No processo de codificação, o quadro é dividido em super macroblocos (FIG. 3.11). Desta forma, um super macrobloco pode ser codificado como dois macroblocos 16x16 na codificação progressiva, ou dois macroblocos 16x16 (um pertencente ao campo par e o outro ao ímpar) na codificação por campo (WANG, 2002a).

Com a introdução desse novo conceito de macrobloco. A codificação por campo pode agora utilizar os mesmo modos presentes na codificação progressiva (FIG. 3.12) (WANG, 2002a).

Abaixo são apresentadas algumas das principais melhorias no estágio de predição que

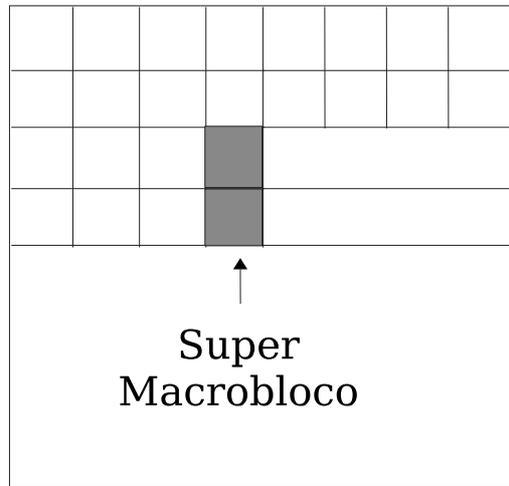


FIG. 3.11: Formação de um Super macrobloco (WANG, 2002a).

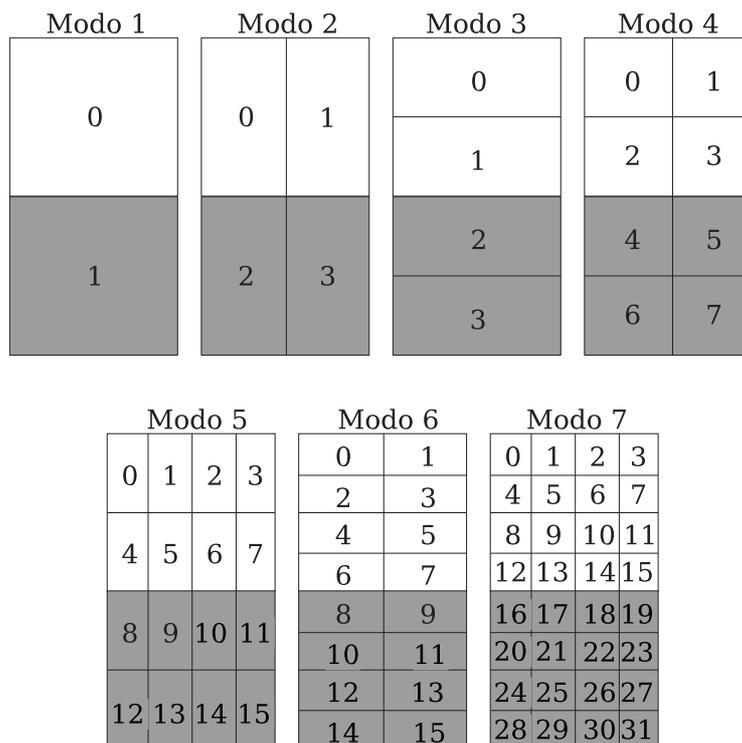


FIG. 3.12: Codificação por Campo - Modos de predição super macrobloco (WANG, 2002a).

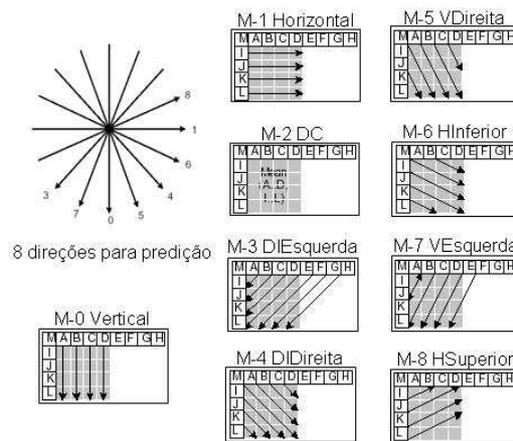
resultam em uma maior acurácia do algoritmo na predição dos macroblocos de entrada. Um melhor detalhamento pode ser encontrado em (RICHARDSON, 2003; WIEGAND, 2003).

- **Modo Intra-Quadro**

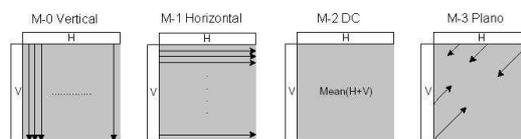
- Na predição intra-quadro, o macrobloco codificado e reconstruído é subtraído do próximo macrobloco que chega à entrada do codificador. As amostras de luminância para a predição podem ser formadas por macroblocos (16x16 amostras) e blocos de 4x4 amostras. As amostras de crominância (Cb e Cr) são formadas por blocos 8x8 amostras. Existem nove modos de predição opcionais para todos os blocos 4x4 de luminância (FIG. 3.13a), quatro para macroblocos 16x16 de luminância (FIG. 3.13b) e quatro para blocos 8x8 crominância. A predição intra-quadro também oferece um codificação *I PCM*, onde os valores das amostras do quadro são transmitidos diretamente (sem predição e transformada) (RICHARDSON, 2003; WIEGAND, 2003).

- **Modo Inter-Quadro.**

- Filtro redutor de blocagem: o codificador H.264/AVC é baseado em blocos. O filtro diminui o efeito de blocos gerado pelo processo de codificação, melhorando o desempenho do processo de estimação no casamento do macrobloco atual dentro da imagem de referência.
- 4 tipos de segmentação do macrobloco para compensação de movimento e 4 tipos de segmentação para partição 8x8.
- Acurácia do vetor de movimento de até  $\frac{1}{4}$  de pixel.
- Vetores de movimento sobre os limites do quadro. Com o uso da técnica de extrapolação dos limites da imagem, os mesmos não ficam mais restritos a apontar para áreas dentro dos quadros previamente decodificados para servir de referência.
- Compensação de movimento com múltiplos quadros de referência: a busca do macrobloco atual é estendida a uma lista de quadros de referência.



(a) Blocos 4x4 (RICHARDSON, 2003; WIEGAND, 2003)



(b) Blocos 16x16 (RICHARDSON, 2003; WIEGAND, 2003)

FIG. 3.13: Modos de predição Intra para blocos 4x4 e macroblocos 16x16 de luminância.

- Ordem dos quadros de referência separada da ordem de exibição dos quadros.
- Quadros tipo B podem ser usados como quadros referência.

No segundo estágio do algoritmo, é aplicada a transformada (T) nos resíduos, sendo os mesmo quantizados (Q). A transformada principal no processo de codificação do H.264/AVC é a DCT, e, diferentemente da maioria dos padrões anteriores, ela é baseada em blocos 4x4. Isto aumenta a acurácia do codificador, diminuindo os artefatos (efeito de bloco) na imagem reconstruída. Abaixo, temos algumas melhorias na transformada dos resíduos dos macroblocos (RICHARDSON, 2003; WIEGAND, 2003).

- Transformada de valores inteiros.
- Tamanho de palavra pequeno: palavras de 16 bits.
- Transformada inversa exata: sem perdas de acurácia.
- 2 tipos de transformadas (DCT e *Hadamard*).

- *Hadamard* Blocos 4x4 para coeficientes DC de luminância para predição de macroblocos no modo intra-quadro 16x16.
  - *Hadamard* Blocos 2x2 para coeficientes DC de crominância (qualquer macro-bloco).
  - DCT - Blocos 4x4 para todos os outros dados residuais.
  - DCT - Blocos 8x8 para todos os outros dados residuais no perfil *High*.
- Tamanho de bloco reduzido.
    - Os padrões anteriores utilizam tamanhos de bloco 8x8 na transformada, enquanto o H.264/AVC utiliza como padrão o tamanho de bloco 4x4. Isto acarreta em uma diminuição dos efeitos de blocos no processo de codificação.
  - Quantização.
    - Não linear.
    - Com 52 valores de parâmetros de quantização (PQ) (O valor do parâmetro de quantização determina o tamanho dos intervalos dos níveis de saída do quantizador).

O último estágio do algoritmo apresenta dois possíveis codificadores por entropia. Os dois são baseados em contexto adaptativo para melhorar o desempenho do codificador, em comparação com os padrões anteriores. O primeiro é denominado de *Context-Adaptive Variable-Length Coding* (CAVLC) e o segundo, mais avançado do que o anterior, é conhecido como *Context-Adaptive Binary Arithmetic Coding* (CABAC) (RICHARDSON, 2003; WIEGAND, 2003).

### 3.5 CODIFICAÇÃO DE VÍDEO ESTÉREO

A viabilização de aplicações que utilizem sistemas de vídeo binocular se depara com alguns obstáculos, tais como: a necessidade de um codificador de vídeo que seja suficientemente robusto, que mantenha a qualidade da seqüência estéreo, rápido e com uma complexidade computacional atraente.

Vários métodos para codificação de vídeo estéreo já foram propostos. Em (YAN, 2003) um método de codificação de vídeo é proposto baseado na interpolação da estimação de movimento e disparidade. Ele pertence aos métodos baseados na intensidade e é chamado de IMDE (*Interpoled Motion and Disparity Estimation*), usando como codificador o MPEG-2 (ITU-T, 1994). Outro método baseado em intensidade encontra-se em (PERKINS, 1992). O mesmo estima o campo de movimento dos quadros consecutivos de um canal de vídeo, ou as disparidades correspondentes entre os pares de quadros estéreos da imagem, por manipulação dos quadros individuais para aumentar o ganho na codificação das seqüências estéreo. Em (YAN, 2003) temos um método baseado na interpolação do quadro estimado. Uma das seqüências é codificada e transmitida conforme o padrão MPEG-2 (ITU-T, 1994), sendo esta denominada de *stream* de referência e a outra de *stream* alvo. No processo de codificação, alguns quadros são descartados e no decodificador uma nova técnica de interpolação e de compensação de quadros estereoscópicos é utilizada para realizar a reconstrução dos quadros descartados. Esse nova técnica é denominada de SFEI-BLCF (*Stereo Frame Estimation Interpolation-Bayesian Least Cost Function*).

Já o método proposto em (TZOVARAS, 1997) baseia-se na manipulação do objeto da cena. O mesmo define as características dos objetos que participam da cena para estimar os campos de disparidade ou de movimento entre os quadros do vídeo estereoscópico, usando um compensador de disparidade e de movimento 3D. Em (LIN, 1999) um sistema baseado na relação taxa-distorção com segmentação *quadtree* para a codificação do vídeo estéreo, é proposto para definir uma técnica de estimação de vetor de deslocamento ótimo para uma qualidade máxima da imagem reconstruída levando em consideração uma determinada taxa de bits. Outro método que utiliza uma decomposição *quadtree* é apresentado em (ELLINAS, 2005).

O método em (PEI, 1998) faz uso do modelo espaço-temporal do SVH e de um estimador de disparidades de correlação binária, suportando baixas taxas de bits para transmissão mantendo uma excelente qualidade visual e tendo uma complexidade computacional baixa.

Quatro métodos de codificação de vídeo estéreo foram utilizados nos testes desta dissertação.

O primeiro método codifica o par de seqüências estéreo, separadamente, como duas seqüências monoculares explorando assim somente as redundâncias espaciais e temporais

da seqüência, sem explorar a redundância binocular presente. Este método é conhecido como *Simulcast* (HASKELL, 1997; TSENG, 1996). (FIG. 3.14a)

Além de ser um método direto na codificação de vídeo binocular, onde as seqüências das vistas esquerdas e direitas são codificadas separadamente, o *Simulcast* aumenta a complexidade no manuseio do trem de bits resultante do processo de codificação, pois o mesmo necessita de sincronização para os quadros na decodificação das seqüências, para exibição e armazenamento.

O segundo e o terceiro método foram propostos em (SUN, 2003), como suporte à codificação de vídeo estéreo do codificador H.264/AVC. Estas propostas têm o intuito de reduzir a complexidade do manuseio do trem de bits, uma vez que tratam uma seqüência binocular dentro de um único trem de bits. Uma alteração no código do software de referência do codificador/decodificador do H.264/AVC foi também proposta, com a finalidade de que o codificador/decodificador pudesse distinguir se a seqüência/trem de bits de entrada seria tratada como estéreo ou monocular, bem como, no caso de ser estéreo, se seria “concatenada” ou “entrelaçada”.

A codificação do método “Entrelaçado” seria realizada usando o modo de codificação de vídeo entrelaçado presente no H.264/AVC. Ainda seria possível prover uma decodificação escalável, onde na codificação entrelaçada o campo par seria codificado somente com quadros de referência formados pelos campos pares anteriores e o campo ímpar poderia fazer uso tanto dos quadros de referência formados pelos campos pares ou ímpares anteriores e do campo par (já codificado) do seu próprio quadro entrelaçado desmontado, garantindo a compatibilidade com sistemas monoculares atuais, pois pelo menos uma das seqüências (esquerda ou direita) poderia ser decodificada sem usar como referência quadros presentes no outro trem de bits ou campo. Para isso, seria preciso uma pequena restrição no codificador. Essa opção é importante para atender a equipamentos sem a funcionalidade de monitores 3D. Esta proposta foi patenteada em abril de 2005 (SHAWMIN, 2005).

No segundo método, os quadros da seqüência esquerda são “concatenados” com os da seqüência direita formando quadros compostos (FIG. 3.14b), onde a primeira metade do quadro composto é formado pelo quadro esquerdo e a segunda pelo direito, apresentando o dobro do número de colunas da seqüência original. Esta nova seqüência monocular é então codificada. Este método foi denominado nesta dissertação como “Concatenado” FIG. 3.14b.

Na FIG. 3.14b, vemos que este método, diferentemente do método *Simulcast*, adapta uma seqüência binocular em uma única seqüência, diminuindo a complexidade no manuseio do trem de bits resultante do processo de codificação, pois o mesmo não requer sincronização para os quadros na decodificação da seqüência. Neste método foram utilizados tamanhos de vetores de movimento que não possibilitam a predição da vista direita pela esquerda. A predição é feita com vistas da mesma seqüência. No decodificador a seqüência de quadros “concatenados” é desmontada gerando as seqüências codificadas das vistas esquerda e direita. Assim como o método anterior, ele explora somente as redundâncias espaciais e temporais da seqüência estéreo.

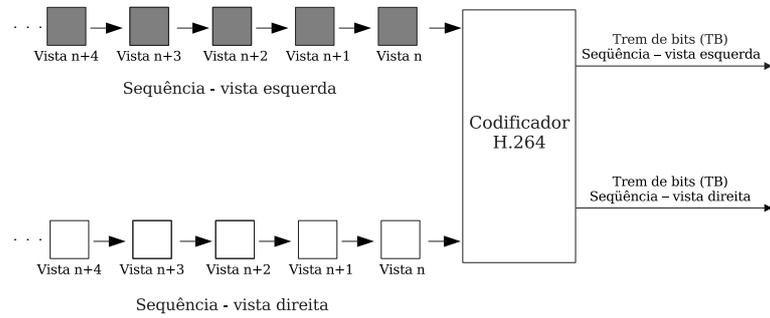
No terceiro método, os quadros da seqüência esquerda são entrelaçados com os da seqüência direita, formando uma nova seqüência monocular para ser codificada, onde cada quadro apresenta o dobro do número de linhas da seqüência original, sendo que a primeira linha corresponde à primeira linha do quadro esquerdo, a segunda linha corresponde à primeira linha do quadro direito, a terceira linha corresponde à segunda linha do quadro esquerdo, e assim sucessivamente. Este método foi denominado nesta dissertação como “Entrelaçado” e é ilustrado pela FIG. 3.14c.

Na FIG. 3.14c, vemos que esse método, assim como o método “Concatenado”, também adapta uma seqüência binocular dentro de uma única seqüência, diminuindo a complexidade no manuseio do trem de bits resultante do processo de codificação, por não necessitar de sincronização para decodificação. No decodificador a seqüência de quadros entrelaçados é desmontada, gerando as seqüências codificadas das vistas esquerda e direita. Assim como os métodos anteriores, este também explora somente as redundâncias espaciais e temporais da seqüências estéreo.

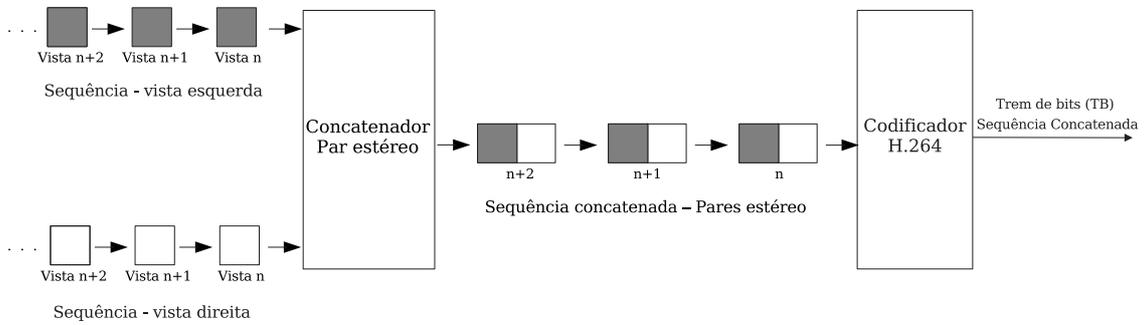
Nas simulações desta dissertação, o codificador H.264/AVC foi utilizado sem nenhuma alteração em seu código. As seqüências “concatenadas” e “entrelaçadas” foram montadas fora do codificador e enviadas para o mesmo para serem codificadas.

Soluções para a codificação de seqüências binoculares que exploram, além das redundâncias espaciais e temporais da seqüência, também suas redundâncias binoculares, fazem uso de conceitos de estereoscopia e disparidade estéreo.

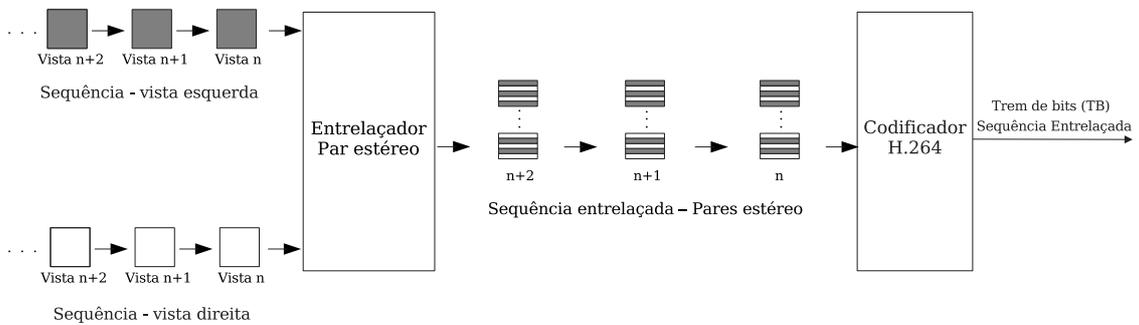
O método quatro utiliza um estimador de disparidades na realização do processo de codificação da seqüência binocular. A FIG. 3.15 mostra o diagrama do processo básico de estimação e compensação de disparidade (esquerdo-direita). As vistas esquerda (VE) e direita (VD) originais são enviadas ao estimador de disparidades (ED), gerando desta forma



(a) Método de codificação *Simulcast*



(b) Método de codificação “Concatenado”



(c) Método de codificação “Entrelaçado”

FIG. 3.14: Métodos de codificação *Simulcast*, “Concatenado” e “Entrelaçado”.

o mapa de disparidades esquerdo-direito (MD). O mapa de disparidades esquerdo-direito e a vista esquerda são codificados (VECOD e MDCOD). No processo de compensação de disparidade, o compensador de disparidade (CD) recebe a vista de referência (VEDECOD = vista esquerda decodificada), a vista direita (VD) e o mapa de disparidades esquerdo-direito (MDDECOD) decodificado (PAGLIARI, 1999). Com a vista esquerda (VEDECOD) e o mapa de disparidades esquerdo-direito (MDDECOD), o compensador de disparidade (CD) é capaz de reconstruir a vista direita ( $VD_{rec}$ ), de acordo com a equação

$$VD_{rec} = \sum_{i=0}^{i=nlinha} \sum_{j=0}^{j=ncol} VEDECOD(i, j + (MDDECOD(i, j))) \quad (3.1)$$

Ainda no compensador, é feita a diferença entre a vista direita original (VD) e a reconstruída ( $VD_{rec}$ ), gerando o resíduo do compensador de disparidade (RCD). Esta operação é dada pela equação

$$RCD = VD - VD_{rec} = \sum_{i=0}^{i=nlinha} \sum_{j=0}^{j=ncol} VD(i, j) - VD_{rec}(i, j) \quad (3.2)$$

O RCD é então enviado ao codificador.

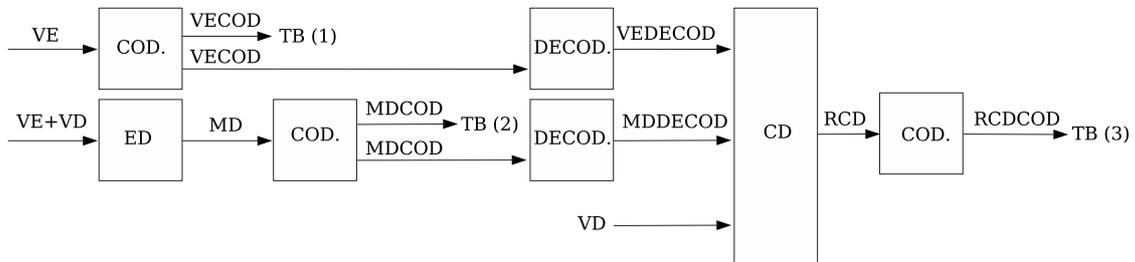


FIG. 3.15: Estimador e compensador de disparidades.

Neste processo são transmitidos ou armazenados três trens de bits, correspondentes à vista esquerda codificada (VECOD), o mapa de disparidades esquerdo-direito (MDCOD) e ao resíduo de compensador de disparidades (RCDCOD). No decodificador, a vista esquerda (VECOD), o mapa de disparidades esquerdo-direito (MDCOD) e o resíduo de compensador de disparidade (RCDCOD) são decodificados e enviados ao compensador de disparidade, onde a vista direita original é reconstruída pela equação

$$VD = \sum_{i=0}^{i=nlinha} \sum_{j=0}^{j=ncol} VEDECOD(i, j + (MDDECOD(i, j))) + RCDDECOD(i, j) \quad (3.3)$$

A precisão na estimação da disparidade é muito importante para diminuir a quantidade de informação transmitida, pois, quanto melhor for o processo de estimação, melhor será a reconstrução da vista direita ( $VD_{rec}$ ) para fazer predição com a vista direita (VD), gerando desta forma um resíduo com muitas regiões homogêneas, a ser enviado ao codificador (PAGLIARI, 1999).

A FIG. 3.16 ilustra a codificação de uma seqüência estéreo com o método quatro. A primeira vista esquerda (VE) é enviada para o estimador de disparidades (ED) juntamente com a vista direita correspondente - VD, gerando o mapa de disparidades estéreo da vista esquerda para a direita (MD). A vista esquerda (VE) e o mapa de disparidades estéreo (MP) são enviados ao codificador H.264/AVC. A vista esquerda codificada (VECOD) e o mapa codificado (MDCOD) são decodificados (VEDECOD e MDDECOD) e enviados ao compensador de disparidade estéreo (CD) juntamente com a vista direita (VD), gerando o resíduo do compensador de disparidade estéreo (RCD), que também é enviado ao codificador H.264/AVC. Na FIG. 3.16, as três seqüências enviadas ao H.264/AVC são codificadas e decodificadas individualmente.

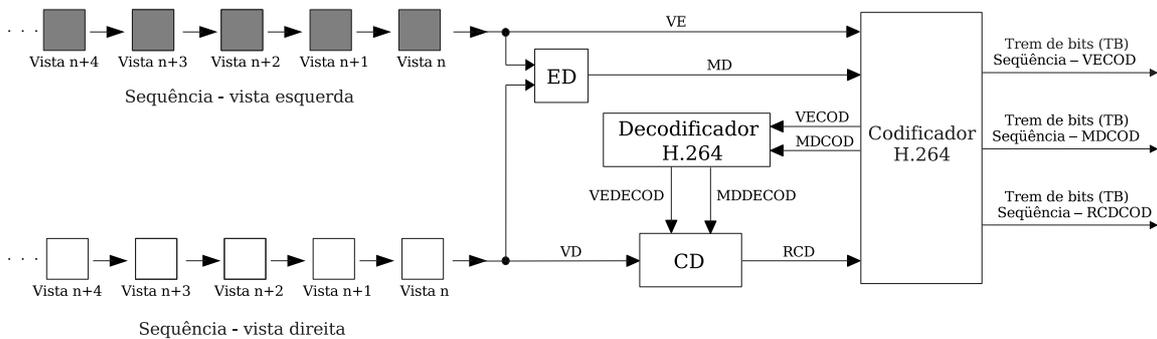


FIG. 3.16: Método de codificação estéreo com estimador de disparidades.

Na FIG. 3.16 vemos que esse método explora a redundância binocular contida em cada par estéreo, mas aumenta a complexidade no manuseio do trem de bits resultante do processo de codificação, pois gera três trens de bits, tendo a necessidade de sincronização no decodificador. Além disso, o método também tem maior complexidade computacional

no processo de codificação, por utilizar um estimador de disparidades, e depende da acurácia do mesmo para alcançar um bom desempenho.

### 3.6 RESUMO

Neste capítulo foram apresentados alguns conceitos básicos de codificação de vídeo monocular e estéreo, e da estrutura e características do codificador H.264/AVC, abordados ou empregados nesta dissertação. Duas ferramentas utilizadas na codificação de vídeo entrelaçado (PAFF e MBAFF), utilizados em conjunto com o terceiro método desta dissertação foram estudadas. Foram apresentados os quatro métodos de codificação de vídeo estéreo utilizados nesta dissertação, sendo que dois deles foram propostos em (SUN, 2003), como suporte à codificação de vídeo estéreo no codificador H.264/AVC. O primeiro método é denominado de *simulcast* e o segundo combina os quadros esquerdo e direito em um único quadro composto, onde a primeira metade do mesmo é formada pelo quadro esquerdo e a segunda pelo direito. O terceiro método entrelaça os quadros esquerdo e direito em um único quadro composto, onde o campo par é formado por todas as linhas do quadro esquerdo e o campo ímpar pelo direito. O quarto método utiliza um estimador de disparidades estéreo para explorar a redundância binocular presente em cada par estéreo, e produz três trens de bits.

O método *Simulcast* e o “Binocular” resultam em dois e três trens de bits, respectivamente, a serem codificados e transmitidos para o decodificador, aumentando a complexidade no manuseio do trem de bits, devido a necessidade de sincronização no processo de decodificação. Já os dois outros métodos diminuem a complexidade no manuseio do trem de bits, pois convertem a seqüência estéreo em uma única seqüência monocular, resultando em um único trem de bits a ser transmitido. Somente o quarto método explora a redundância binocular das seqüências estéreo.

## 4 TESTE DO DESEMPENHO DOS MÉTODOS

Neste capítulo é apresentado o ambiente de testes utilizado nas simulações, os critérios de avaliação, bem como a comparação das seqüências utilizadas e os resultados obtidos com os quatro métodos de codificação de vídeo estéreo usando o codificador H.264/AVC. Na Seção 4.1 é feita uma breve introdução sobre os resultados obtidos. O ambiente de testes é descrito na Seção 4.2. Na Seção 4.3 é apresentada a justificativa do ganho de codificação presente no método “Entrelaçado”. Na Seção 4.4 são apresentados os resultados dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” para taxas de bits constantes (TBC). Os resultados obtidos para taxas de bits variáveis (TBV) encontram-se na Seção 4.5 e na Seção 4.6 temos os resultados com o método “Binocular” em comparação com o método *Simulcast*. Na Seção 4.7 apresenta-se o resumo do capítulo.

### 4.1 INTRODUÇÃO

A avaliação do desempenho dos quatro métodos de codificação de vídeo estéreo usando o codificador H.264/AVC mostra que o primeiro (*Simulcast*), em alguns casos, utiliza mais bits no processo de codificação, comparado com o segundo (“Concatenado”) e o terceiro (“Entrelaçado”) métodos. O método “Entrelaçado” apresentou o melhor desempenho entre os quatro métodos avaliados. Uma das razões desse ganho é a possibilidade de realizar a predição da vista direita (campo ímpar) com a própria vista esquerda (campo par codificado) do mesmo quadro “entrelaçado” desmontado. O quarto método (“Binocular”) tem seu desempenho diretamente associado à qualidade do mapa gerado no processo de estimação de disparidade. A título ilustrativo, foi mostrado que uma vez que o estimador de disparidades obtenha um mapa preciso, o método “Binocular” pode alcançar um desempenho superior aos demais métodos avaliados. A avaliação da complexidade dos quatro métodos de codificação de vídeo estéreo mostra que o primeiro (*Simulcast*) e o último (“Binocular”) geram dois, e três trens de bits, tendo a necessidade de sincronização dos quadros no processo de decodificação das seqüências.

No método “Entrelaçado” é realizada uma codificação por campo da seqüência de

vídeo estéreo entrelaçada. Quando for utilizada alguma ferramenta de codificação de vídeo entrelaçado presente no codificador H.264/AVC, será feita uma referência à mesma.

Seqüências estéreo com diferentes graus de movimento e textura são avaliadas nesta tese. O ambiente de teste utilizado nas simulações para a avaliação do desempenho dos quatro métodos é descrito com mais detalhes na Seção 4.2.

## 4.2 AMBIENTE DE TESTE

A versão utilizada do codificador H.264/AVC para as simulações desta dissertação foi a JM-9.4 (JVT). Os perfis e níveis utilizados nas seqüências foram (*Main* e *High*) e (3.0).

Todos as simulações usaram cinco quadros de referências na estimação de movimento, intervalo de busca igual  $\pm 32$ , estimação de movimento rápida desligada, transformada de *Hadamard* ligada e *RDOptimization* (taxa-distorção) ligado.

Nas simulações foram utilizadas três estruturas de GoP conforme abaixo:

- GoP I - O primeiro quadro é do tipo I (intra) e os quadros restantes da seqüência são do tipo P (preditos do quadro anterior).
- GoP II - O tamanho da estrutura do GoP é 15 (N=15). A mesma possui M=3 (distância entre quadros I-P, P-P ou P-I) sendo formada por quadros I, P e B).
- GoP III - O tamanho da estrutura do GoP é 12 (N=12). A mesma possui M=2 (distância entre quadros I-P, P-P ou P-I) sendo formada por quadros I, P e B).
- Os quadros entre parênteses são codificados como dois campos. Desta forma, quadros tipo I podem ser codificados como um quadro tipo I ou dois campos tipo I ou o primeiro como I e o segundo como P. Quadros tipo P podem ser codificados como um quadro tipo P ou dois campos tipo P e os quadros tipo B podem ser codificados como um quadro tipo B ou dois campos tipo B.

– Codificação por quadro.

\* GoP I: I, P, P, P ...,

\* GoP II: I, B, B, P, B, B, P ...,

\* GoP III: I, B, P, B, P, B, P ...,

– Codificação por campo.

- \* GoP I: (II ou IP), (PP), (PP), (PP) ...,
  - \* GoP II: (II ou IP), (BB), (BB), (PP), (BB), (BB), (PP) ...,
  - \* GoP III: (II ou IP), (BB), (PP), (BB), (PP), (BB), (PP) ...,
- Codificação adaptativa quadro/campo.
- \* GoP I: I ou (II ou IP), P ou (PP), P ou (PP), P ou (PP) ...,
  - \* GoP II: I ou (II ou IP), B ou (BB), B ou (BB), P ou (PP), B ou (BB), B ou (BB), P ou (PP) ...,
  - \* GoP III: I ou (II ou IP), B ou (BB), P ou (PP), B ou (BB), P ou (PP), B ou (BB), P ou (PP) ...,

#### 4.2.1 SEQÜÊNCIAS UTILIZADAS

Nas simulações dos métodos foram utilizadas quatro seqüências de vídeo estéreo em tons de cinza (MAN (HHI)<sup>1</sup>, SAXO (CCETT)<sup>2</sup>, INDOOR, OUTDOOR (LABVISIONE)<sup>3</sup> e RANDOT (MCL)<sup>4</sup>). MAN (384x384) é uma seqüência estéreo do tipo *head and shoulders* (o enquadramento da cena tem seu ponto central na região entre o tórax e a cabeça da pessoa) típica de vídeo conferência, com muitas áreas de baixíssima textura e com variação de movimento inter-quadro de baixa a moderada. Cada par estéreo apresenta altos valores de disparidade. SAXO (352x288) é uma seqüência estéreo que apresenta partes com detalhes (camisa do saxofonista) e com baixa textura (área dentro de alguns objetos), com uma variação de movimento inter-quadro moderada. Os pares estéreo apresentam valores de disparidade de moderados a baixos. INDOOR e OUTDOOR (352x288) são seqüências estéreo que apresentam grandes áreas de baixa textura, com variação de movimento inter-quadro de baixa a alta, e valores de disparidade pequenos. A seqüência RANDOT foi gerada artificialmente através do par estéreo sintético RANDOT. Ela apresenta mapas com disparidades conhecidas para todos os pares estéreo da seqüência.

As seqüências SAXO e MAN possuem 51 quadros, INDOOR e OUTDOOR 100 quadros e a RANDOT 50 quadros. As seqüências obtidas para as simulações apresentam um

---

<sup>1</sup>Heinrich-Hertz-Institut

<sup>2</sup>Centre Commun d'Études de Télédiffusion et Télécommunications

<sup>3</sup>Labvisione

<sup>4</sup>Multimedia Communications Laboratory

número pequeno de quadros. Seqüências maiores não foram utilizadas, pois, o número de seqüências estéreo disponíveis é pequeno com essa característica.

#### 4.2.2 MEDIDAS DE AVALIAÇÃO

Na comparação do desempenho de métodos de codificação de vídeo estéreo disponíveis na literatura, são levados em consideração os seguintes critérios (SETHURAMAN, 1994; SULLIVAN, 2004).

- **Taxa de compressão:** valor de compressão alcançado por um método para uma determinada seqüência de entrada do codificador, sendo a razão entre a taxa de bits por pixels (bpp) da seqüência de entrada pela taxa de bits por pixels (bpp) da mesma codificada. (SETHURAMAN, 1994).
- **Fator uso de banda:** razão entre a taxa de bits por pixels (bpp) utilizada pelo método *Simulcast* e a taxa de bits por pixels (bpp) utilizada pelos demais métodos simulados para uma determinada seqüência de entrada do codificador (SETHURAMAN, 1994).
- **Métrica de Qualidade:** relação de similaridade entre a seqüência de entrada e a seqüência decodificada. A medida utilizada para esta comparação é a PSNR (*Peak Signal-to-Noise Ratio*) (SULLIVAN, 2004) (nos testes foi considerado somente a componente de luminância (PSNR-Y)).

O método *Simulcast*, por ser o método direto na codificação de vídeo estéreo, foi escolhido nesta dissertação como o método de referência.

Na EQ. 4.1 temos a relação utilizada para medir a taxa de compressão , e na EQ. 4.2 o fator de uso de banda. A medida utilizada para determinar a qualidade da seqüência reconstruída é o PSNR (EQ. 4.3), somente na componente de luminância (PSNR-Y), pois as imagens e seqüências utilizadas são em escala de cinza (8 bits por pixels). Esta medida é freqüentemente utilizada na determinação da qualidade da seqüência codificada reconstruída (SULLIVAN, 2004). No método *Simulcast* (I) o cálculo do PSNR-Y é feito diretamente sobre as seqüências codificadas reconstruídas esquerda e direita. Nos métodos “Concatenado” (II) e “Entrelaçado” (III), o cálculo do PSNR-Y foi feito utilizando as

seqüências esquerda e direita reconstruídas, que foram extraídas das seqüências “concatenadas” e “entrelaçadas” reconstruídas. No método “Binocular” (IV), o cálculo para as seqüências esquerda e direita reconstruídas foi feito de modo idêntico ao método *Simulcast* (I), sendo a seqüência direita reconstruída gerada pelo compensador de disparidades, através da seqüência esquerda, do mapa de disparidades esquerdo-direito e do RCD.

$$TC = \frac{\text{bpp seqüência original } x}{\text{bpp seqüência comprimida } x} \quad x \in \{I, II, III, IV\} \quad (4.1)$$

$$FUB = \frac{\text{bpp seqüências comprimidas método I}}{\text{bpp seqüências comprimidas método II ou III ou IV}} \quad (4.2)$$

$$PSNR(dB) = 10 \log_{10}(u_{max}^2 / EMQ) \quad (4.3)$$

$$EMQ = \left( \frac{1}{nlin * ncol} \right) \sum_{x=1}^{x=nlinha} \left( \sum_{y=0}^{y=ncol} [u(x, y) - \hat{u}(x, y)]^2 \right) \quad (4.4)$$

Na EQ. 4.3,  $u_{max}$  é o valor máximo em níveis de intensidade em uma escala de cinza (neste caso 255 níveis = 8 bits) que cada pixel pode assumir. O erro médio quadrático (EMQ) é dado pela EQ. 4.4, onde  $nlin$  e  $ncol$  são os números de linhas e colunas do quadro,  $u(x, y) - \hat{u}(x, y)$  corresponde à diferença entre um ponto da imagem original e seu correspondente (mesma coordenada) na imagem reconstruída.

#### 4.3 GANHO DE CODIFICAÇÃO UTILIZANDO O MÉTODO “ENTRELAÇADO”

Segundo os proponentes dos métodos “Concatenado” e “Entrelaçado” (SUN, 2003), o ganho na codificação utilizando o método “Entrelaçado” é obtido principalmente da predição das vistas esquerda e direita do quadro entrelaçado. De acordo com os autores, após a primeira vista do quadro entrelaçado ser codificada, a segunda poderia usa-lá como vista de referência na predição entre as vistas.

Em (SUN, 2005), os autores compararam o desempenho do método “Entrelaçado” utilizando as ferramentas de codificação de vídeo entrelaçado PAFF e MBAFF do codificador H.264/AVC. Foram analisadas três configurações: a primeira delas é o método

“Concatenado” presente na proposta em (SUN, 2003) e avaliado neste dissertação. A segunda usa o método “Entrelaçado” com as ferramentas PAFF e MBAFF que possibilitam uma codificação adaptativa da seqüência entrelaçada entre progressiva ou por campo à nível de macrobloco, e a terceira usa também o método “Entrelaçado” com a restrição do campo par não utilizar o campo ímpar como campo de referência na predição entre campos, mas podendo o campo ímpar utilizar tanto os campos pares ou ímpares para sua predição entre campos. Em todas três configurações foram utilizados cinco quadros de referência na estimação de movimento. Os resultados obtidos mostraram que as duas últimas configurações obtiveram um desempenho similar e superior a primeira.

O estudo realizado nesta dissertação demonstra que o grau de distorção perspectiva e o número de quadros de referência utilizados no processo de estimação de movimento são dados importantes no ganho de codificação utilizando o método “Entrelaçado”. Uma das contribuições do codificador H.264/AVC é o uso de um número maior de quadros referência. A codificação de seqüências monoculares se beneficia desta característica deste codificador.

As seqüências utilizadas no artigo (SUN, 2005) não são referenciadas. Através de uma inspeção visual, verifica-se que os pares estéreo usados em (SUN, 2005) apresentam valores de disparidades pequenos <sup>5</sup> (FIG. 4.1). Tais valores e o número de quadros de utilizados como referência na estimação de movimento justificam o ganho na codificação utilizando o método “Entrelaçado” descrito pelos autores. No entanto, quando se utiliza seqüências estéreo com valores de disparidades médios ou altos, o ganho na codificação utilizando o método “Entrelaçado” não é obtido principalmente da utilização das vistas esquerdas codificadas como referência para a predição das vistas direitas do quadro entrelaçado.

Nesta dissertação foram testadas seis configurações com o método “Entrelaçado” utilizando as seqüências MAN e SAXO para demonstrar a relevância da distorção perspectiva em relação ao número de quadros de referência no ganho de codificação do método “Entrelaçado”. A primeira configuração consiste no método “Entrelaçado” analisado nesta dissertação. Na segunda, o ferramenta de codificação de vídeo entrelaçado PAFF é utilizada no método “Entrelaçado”, e na terceira, as ferramentas de codificação de vídeo entrelaçado PAFF e MBAFF são utilizadas. Todos as três configurações usaram o perfil *High*, com GoP II, PQ fixo para a codificação dos quadros do tipo I e P (PQ MAN = 12, 21, 26, 31, 36 e PQ SAXO = 21, 26, 31) e para a codificação dos quadros do tipo B

---

<sup>5</sup>Estas seqüências foram requisitadas para os autores, mas não foram disponibilizadas



(a) Sequências de vídeo estéreo (SUN, 2005) (b) Sequências de vídeo estéreo (SUN, 2005)



(c) Sequências de vídeo estéreo (SUN, 2005)

FIG. 4.1: Sequências de vídeo estéreo “Concatenadas”

(PQ MAN = PQ SAXO = 30), cinco quadros de referências na estimação de movimento, intervalo de busca igual  $\pm 32$ , estimação de movimento rápida desligada, transformada de Hadamard ligada e RDOptimization ligado. As configurações quatro, cinco e seis são as configurações um, dois e três respectivamente, sendo que somente um quadro pode ser utilizado como referência na estimação de movimento.

Foi observado que, para seqüências estéreo com valores de disparidades médios e altos (FIG. 4.2a e FIG. 4.2b), o método não optava, na maioria dos testes, pelo uso da vista esquerda codificada como referência na predição da vista direita do próprio quadro entrelaçado quando foi utilizado um único quadro de referência. Os testes indicam que, quando a seqüência estéreo possui uma disparidade grande (FIG. 4.2a), a predição da vista direita é realizada tendo como referência às vistas direitas codificadas e a predição da vista esquerda às vistas esquerda codificadas dos quadros anteriores e/ou posteriores.

Por isso, quando utilizamos um número de quadros de referência grande (por exemplo: cinco quadros), o desempenho para uma seqüência com valor de disparidade elevado alcançava resultados similares para as duas configurações testadas usando o método “Entrelaçado” em (SUN, 2005) e as três primeiras (método “Entrelaçado”, “Entrelaçado” com a ferramenta PAFF e “Entrelaçado” com as ferramentas PAFF mais MBAFF) usadas nesta dissertação no teste. Uma conclusão mais precisa não foi possível devido à disponibilidade de uma única seqüência estéreo com esta característica.



(a) Quadros esquerdo e direito 25 - MAN



(b) Quadros esquerdo e direito 25 - SAXO



(c) Quadros esquerdo e direito 25 - MAN



(d) Quadros esquerdo e direito 25 - SAXO

FIG. 4.2: Seqüências de vídeo estéreo “Concatenada” e “Entrelaçada”.

Nas FIG. 4.3 e FIG. 4.4, temos as duas seqüências estéreo que foram analisadas para um número de quadros de referências igual a um e cinco. As seqüências foram simuladas nos modos “Entrelaçado”, “Entrelaçado” com a ferramenta PAFF e “Entrelaçado” com as ferramentas PAFF e a MBAFF.

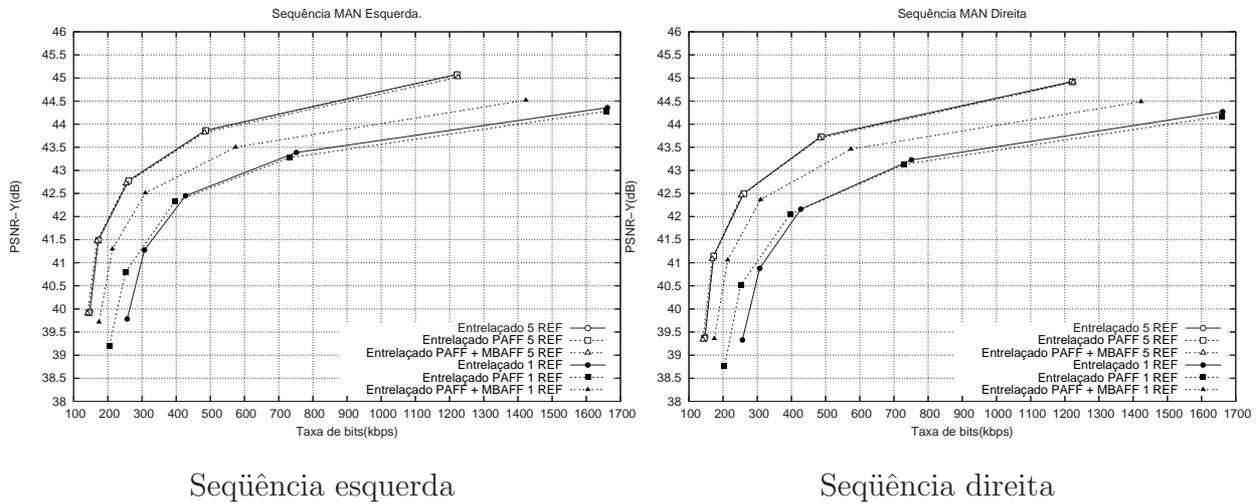


FIG. 4.3: Comparação do Método “Entrelaçado”. Seqüências estéreo MAN.

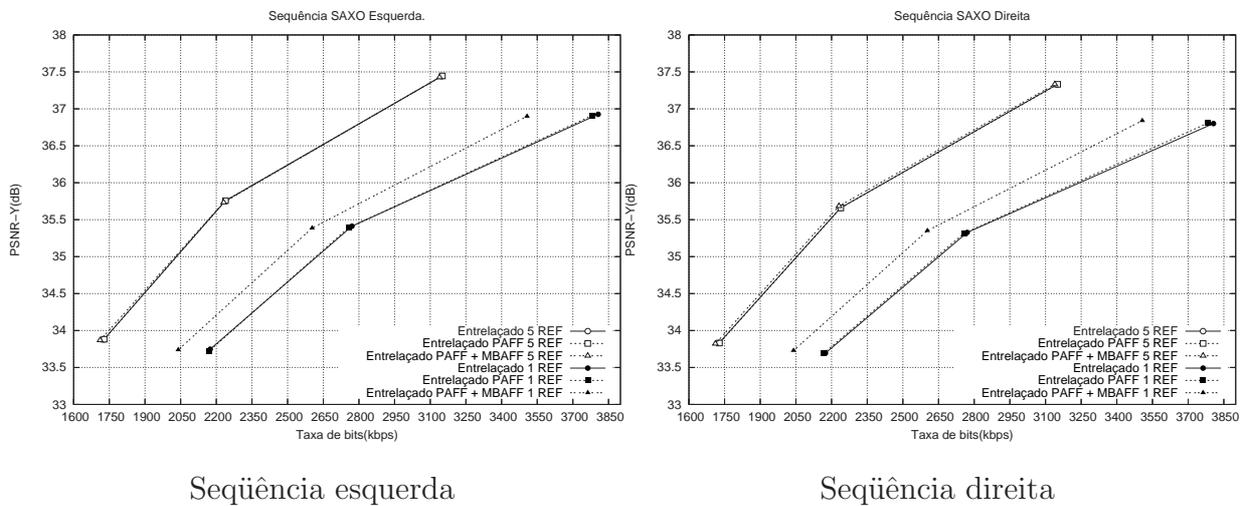


FIG. 4.4: Comparação do Método “Entrelaçado”. Seqüências estéreo SAXO.

Podemos observar que, nas duas seqüências, o desempenho do método “Entrelaçado”, nos três modos, quando são utilizados cinco quadros como referência, é similar (FIG. 4.3 e FIG. 4.4).

Porém, quando o número de quadros de referência é igual a 1, o desempenho do método

“Entrelaçado” apresenta-se inferior, principalmente em relação ao método “Entrelaçado” com as ferramentas PAFF e MBAFF (FIG. 4.3 e FIG. 4.4).

Na seqüência MAN, quando foi utilizado um único quadro de referência, pudemos observar através da *flag DisplayEncParams* que quando habilitada apresenta todos os parâmetros do processo de codificação para ser visualizado ao armazenado, que o número de quadros codificados como quadro progressivo foi na média 6 vezes maior do que quando se utilizou cinco quadros de referência. Estes testes indicam que o aumento do número de quadros de referência melhorou o desempenho do método “Entrelaçado” em seqüências com valores grandes de disparidades, pois as mesmas podem utilizar na predição as vistas dos quadros anteriores e/ou posteriores. No entanto, seria necessária a realização de um número maior de testes com outras seqüências estéreo com altos valores de distorção perspectiva, uma vez que a única seqüência utilizada com esta característica foi a seqüência MAN. Na seqüência SAXO, quando foi utilizado um único quadro de referência, pudemos observar de maneira análoga à seqüência MAN, que o número de quadros codificados como quadro progressivo foi muito próximo do valor utilizado com cinco quadros de referência. A diferença máxima encontrada para um mesmo parâmetro de quantização foi em média de 5 quadros codificados como progressivo.

#### 4.4 COMPARAÇÃO ENTRE OS MÉTODOS - TAXA DE BITS CONSTANTE

O método “Entrelaçado” foi codificado através de duas formas distintas. A diferença entre as duas formas está na utilização da ferramenta PAFF, a qual habilita a codificação adaptativa entre quadro ou campo no codificador H.264/AVC, de forma que o algoritmo escolhe se irá tratar o quadro entrelaçado como um quadro progressivo ou um quadro formado por dois campos.

As estruturas de GoP I (IPPP...) e GoP II (IBBPBBP...) foram utilizadas nas simulações dos métodos.

As quatro seqüências foram codificadas com taxas de bits constantes de valores 100, 200, 300, 400, 500, 1000 e 1500 Kbps, com o intuito de analisar a qualidade da seqüência reconstruída para taxa baixas.

Os parâmetros de quantização (PQ) inicialmente utilizados nos três métodos nos perfis *Main* e *High* para a codificação dos quadros tipo I e P foram {14, 16, 19, 23, 28, 30} para as seqüências MAN, INDOOR e OUTDOOR e {16, 19, 22, 25, 28, 39} para SAXO. O quadro

tipo B foi codificado com um parâmetro de quantização fixo de valor 30.

Esta seção apresenta os resultados das simulações dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” para as seqüências MAN, SAXO, INDOOR e OUTDOOR.

O método “Entrelaçado” apresenta um desempenho superior aos demais métodos na maioria das simulações. Em relação ao método “Concatenado”, além da explicação dada na Seção 4.3, onde o campo ímpar (vista direita) pode utilizar o campo esquerdo (vista esquerda) codificado na predição entre os campos, isso se deve ao fato de que o quadro composto deste método pode apresentar, na junção dos quadros esquerdos e direitos, uma decorrelação entre os pixels vizinhos, pois os mesmos representam regiões diferentes da imagem. Desta forma, os pixels vizinhos ou próximos a junção, referentes a vista direita (lado direito do quadro composto) não representam um deslocamento dos pixels vizinhos ou próximos da junção, referente a vista esquerda (lado esquerdo do quadro composto). Com isso, há uma codificação ineficiente nesta parte do quadro composto. Em relação ao método *Simulcast*, os dois métodos anteriores apresentam um desempenho melhor ou similar, por não terem *overhead*, uma vez que os mesmos adaptam uma seqüência binocular em uma seqüência monocular.

Quando se quer codificar os dois campos de um quadro entrelaçado do tipo I como Intra é necessário habilitar a *flag IntraBottom* no codificador H.264/AVC. Entretanto, quando se utiliza taxa de bits constante, esta *flag* apresenta um erro no processo de codificação gerando uma taxa de bits muito elevada para a seqüência codificada. Desta forma, esta *flag* não foi habilitada nas simulações com taxa de bits constantes. Nas simulações do método “Entrelaçado”, as seqüências reconstruídas da vista direita (campo ímpar) apresentam um desempenho inferior ao das vistas esquerdas (campo par). Esse desempenho inferior é devido ao fato de que quando se codifica um quadro do tipo I, somente o campo par é codificado como Intra.

#### 4.4.1 RESULTADOS DA SEQÜÊNCIA MAN COM O PERFIL *MAIN* E *HIGH*.

Conforme mostrado nesta seção foram utilizados inicialmente alguns parâmetros de quantização na codificação da seqüência MAN. O valor do parâmetro de quantização escolhido para a comparação entre os métodos foi o 23 por apresentar na média a menor variação da qualidade (PSNR-Y) ou a maior correlação (similaridade) entre os quadros codificados com relação as taxas constantes pré-definidas.

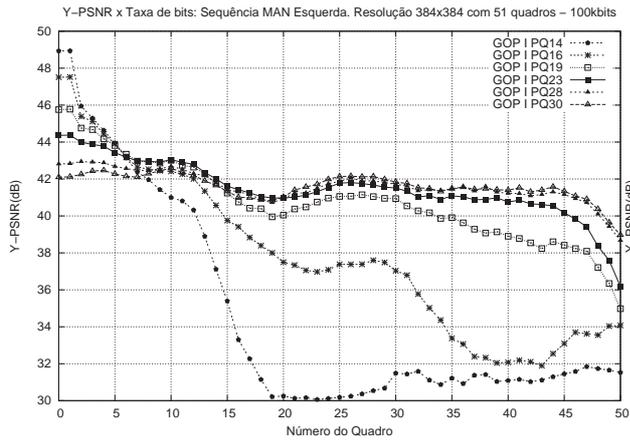
Na FIG. 4.5 temos as curvas de PSNR-Y x quadros da seqüência MAN codificada através do método *Simulcast* GoP I para as taxas de 100, 200, 300, 400 e 500 Kbps. Podemos observar que o parâmetro de quantização de valor 23 apresenta nas taxas de 200, 300 e 400 Kbps, tanto na seqüência esquerda quanto na direita, a menor variação da qualidade (PSNR-Y) ou a maior correlação entre os quadros codificados, ficando ainda entre os três melhores parâmetros com estas características, nas demais taxas utilizadas.

Nas FIG. 4.6a e FIG. 4.6b temos as curvas de PSNR-Y x taxa de bits (curvas taxa-distorção) da seqüência MAN codificada através do método “Entrelaçado” sem e com a ferramenta PAFF utilizando o GoP I.

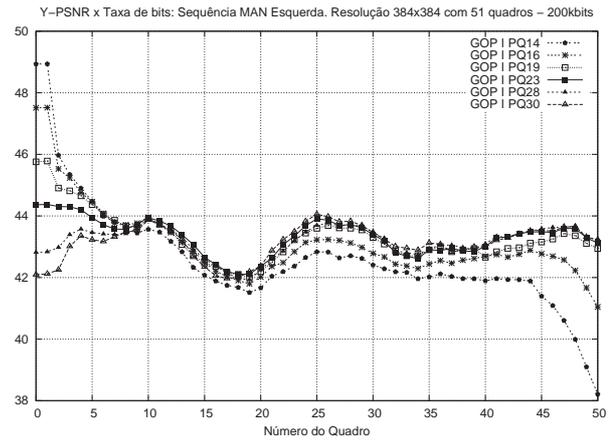
Podemos observar na FIG. 4.6a que a codificação com a ferramenta PAFF apresenta um desempenho superior para taxas de codificação abaixo de 800 Kbps. Na mesma figura, vemos que o método “Entrelaçado” sem a ferramenta PAFF apresenta um ganho a partir de 800 Kbps, devido ao fato do codificador variar o parâmetro de quantização inicial para ajustar a taxa pré-selecionada. O método “Entrelaçado” sem a ferramenta PAFF apresentou uma variação mais suave, com valores de PQ próximos em relação ao método “Entrelaçado” com a ferramenta PAFF. Já na FIG. 4.6b vemos que o método “Entrelaçado” sem a ferramenta PAFF apresenta resultados superiores para taxas de codificação acima de 200 Kbps e resultados similares abaixo desta taxa. Isto se deve ao fato do codificador variar o parâmetro de quantização inicial para ajustar a taxa pré-selecionada. Na FIG. 4.6a, para um valor de PSNR-Y igual a 47,5 dB, o método “Entrelaçado” sem a ferramenta PAFF apresenta uma taxa de aproximadamente 110 Kbps inferior ao método com a ferramenta PAFF, e na FIG. 4.6b, para um valor de PSNR-Y de 46,7 dB, a diferença é de 250 Kbps. A diferença apresentada de aproximadamente 1,2 dB, entre a seqüência reconstruída da vista esquerda e a direita, é devido ao fato da *flag IntraBotton* não estar habilitada.

Nas FIG. 4.7a e 4.7b temos as curvas de PSNR-Y x taxa de bits da seqüência MAN codificada através dos métodos *Simulcast*, “Concatenado” e “Entrelaçado”, sendo o último sem a ferramenta PAFF. Nos gráficos, o eixo horizontal representa para os métodos “Concatenado” e “Entrelaçado” a taxa de codificação binocular das seqüências, uma vez que estes dois métodos convertem a seqüência binocular dentro de uma nova seqüência formada monocular. Para o método *Simulcast*, o eixo horizontal representa a taxa de codificação monocular das seqüências esquerda e direita.

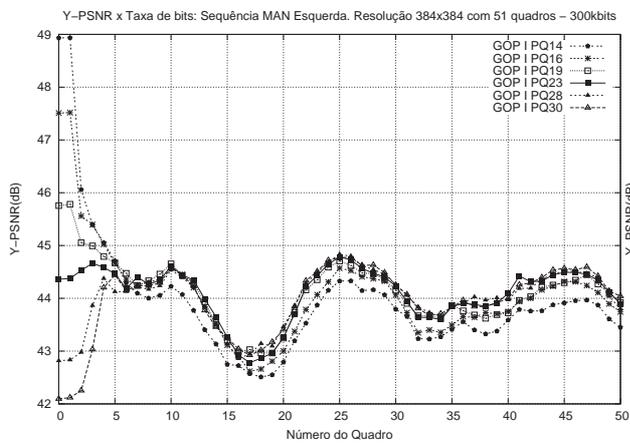
Nas FIG. 4.7a e FIG. 4.7b vemos que o método “Entrelaçado” apresenta um de-



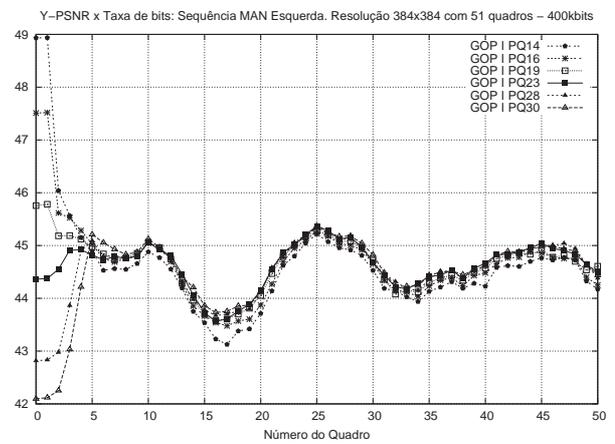
(a) 100 Kbps GoP I PQ23.



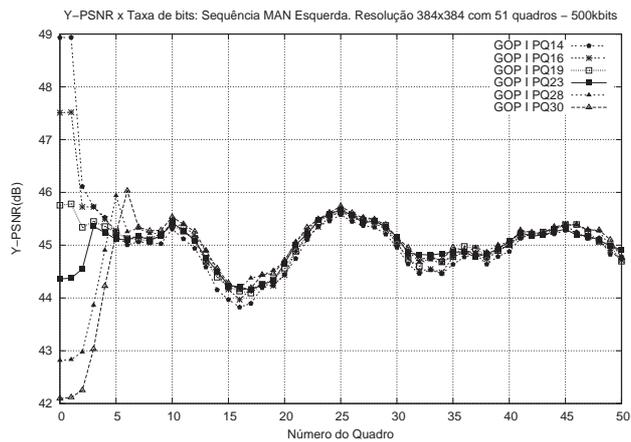
(b) 200 Kbps GoP I PQ23.



(c) 300 Kbps GoP I PQ23.

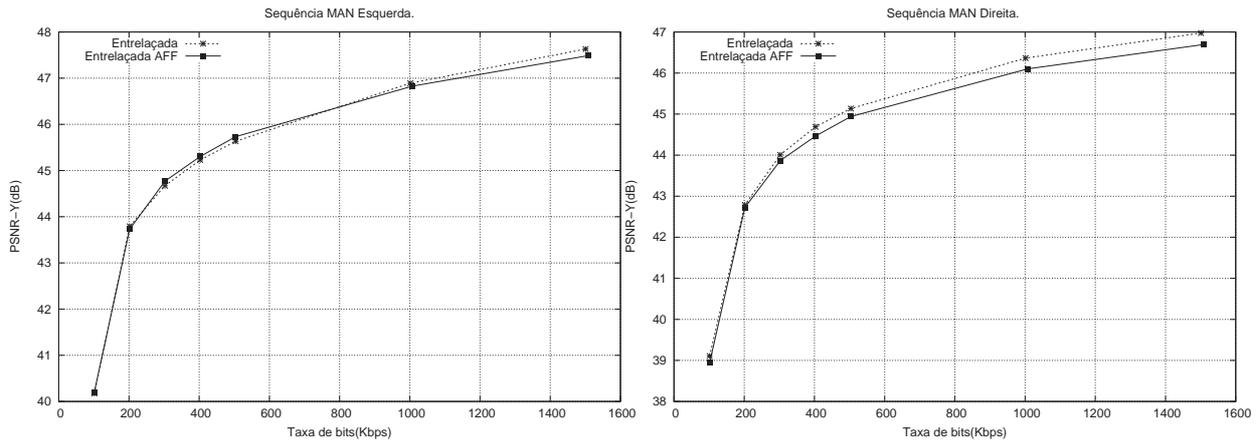


(d) 400 Kbps GoP I PQ23.



(e) 500 Kbps GoP I PQ23.

FIG. 4.5: Comparação PQ - Sequência MAN esquerda *Simulcast* - Perfil *Main* e *High*.



(a) Sequência MAN esquerda.

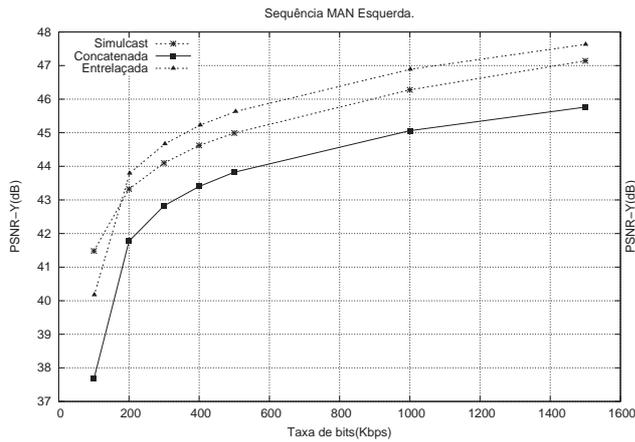
(b) Sequência MAN direita.

FIG. 4.6: Comparação do Método “Entrelaçado” com/sem PAFF. GoP I PQ23 - Perfil *Main* - MAN.

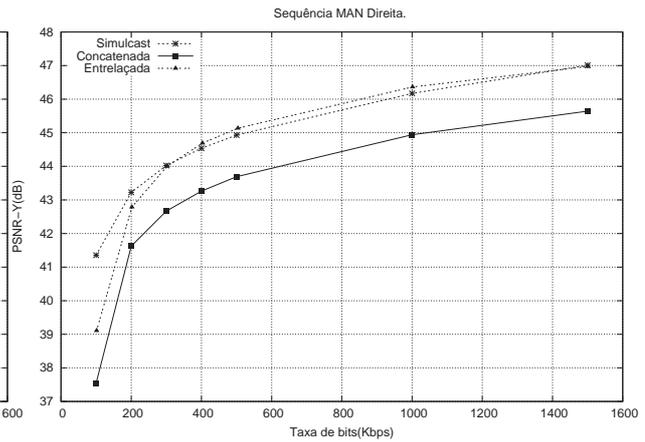
sempenho superior aos métodos *Simulcast* e “Concatenado”. Notamos diretamente que o método Entrelaçado obtém um ganho máximo para uma mesma taxa de codificação sobre o método “Concatenado” nas seqüências esquerda e direita, respectivamente de aproximadamente 1,8 dB e 1,4 dB, para taxas baixas como 100 Kbps, e de aproximadamente 2,5 dB e 1,5 dB para taxas altas como 1500 Kbps. Observamos também que a qualidade das seqüências direitas reconstruídas é inferior quando comparadas com as seqüências esquerdas, pois no processo de codificação, quando o quadro é codificado como Intra, somente o campo par é codificado como Intra, sendo o campo ímpar codificado com predição Inter. Esse problema se repetirá em todas as simulações utilizando o método “Entrelaçado”, pois a *flag IntraBotton*, que quando habilitada força a codificação do campo ímpar ser no modo Intra, estava resultando em uma taxa de codificação extremamente superior à codificação da seqüência sem esta *flag* habilitada.

As FIG. 4.7c e FIG. 4.7d mostram as curvas de PSNR-Y x taxa de bits de 200, 400 e 1000 Kbps da seqüência MAN codificada através dos métodos *Simulcast*, “Concatenado” e “Entrelaçado”, sendo o último sem a ferramenta PAFF. O eixo horizontal do gráfico é relativo à taxa de codificação da seqüência binocular (vistas esquerda e direita). Desta maneira, temos como comparar de uma forma direta os métodos “Concatenado” e “Entrelaçado” com o método *Simulcast*.

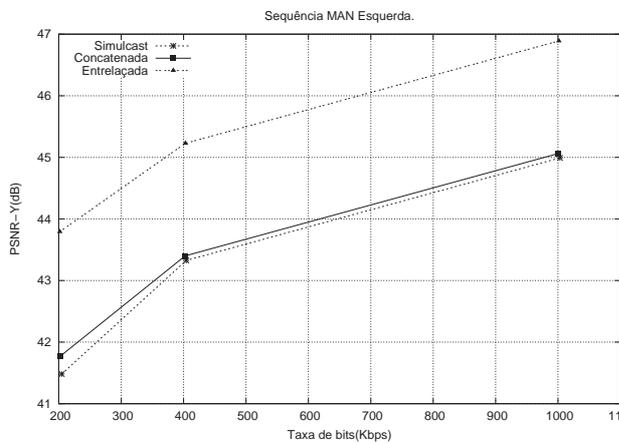
Vemos que o método “Entrelaçado” apresenta um desempenho superior ao método *Simulcast*. Ele obtém um ganho máximo para uma mesma taxa de codificação sobre



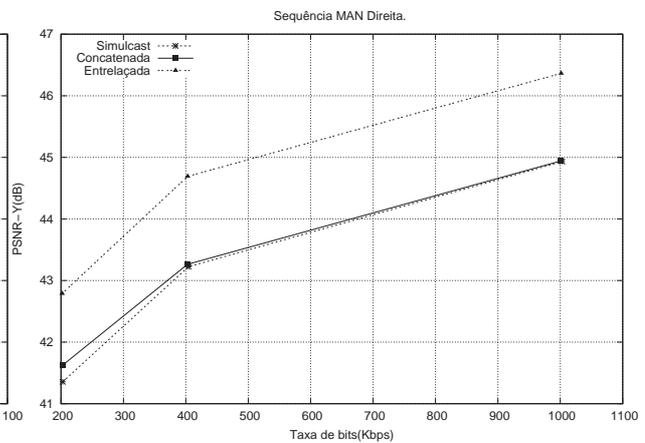
(a) Sequência MAN esquerda.



(b) Sequência MAN direita.



(c) Sequência MAN esquerda.



(d) Sequência MAN direita.

FIG. 4.7: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” sem PAFF. Sequência MAN - GoP I PQ23 - Perfil *Main*.

o método *Simulcast* de aproximadamente 2,4 dB para taxas baixas como 200 Kbps, e de aproximadamente 2,0 dB para taxas altas como 1000 Kbps. Para um mesmo valor de qualidade (PSNR-Y), como por exemplo 45,0 dB, na FIG. 4.7c temos uma economia na taxa de codificação com método “Entrelaçado” em relação aos outros dois métodos de aproximadamente 625 Kbps. Para a seqüência direita (FIG. 4.7d) a economia é de 550 Kbps para um valor de PSNR de aproximadamente 44,7 dB. Observamos também que o método “Concatenado” apresenta um desempenho melhor ou similar ao método *Simulcast* obtendo um ganho de até 0,3 dB (200 Kbps) nas FIG. 4.7c e FIG. 4.7d. Analisando o arquivo de dados estatísticos do processo de codificação gerado pelo codificador H.264/AVC, vemos que o método “Concatenado” apresenta uma economia de bits no

cabeçario do trem de bits e na estimação de movimento. Sendo esta, uma justificativa para o seu desempenho similar e em algumas taxas superior ao método *Simulcast*.

A seqüência MAN também foi simulada no perfil *Main* com GoP II e no perfil *High* com GoP I e II. O comportamento das curvas dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” nestas simulações são similares ao descrito para as curvas com o perfil *Main* com GoP I. Desta forma, as justificativas de ganho de desempenho dos métodos “Entrelaçado” e “Concatenado” apresentados anteriormente para o GoP I, também foram utilizadas na análise destas curvas.

Os resultados do método “Entrelaçado” com e sem a ferramenta PAFF habilitada foram similares. Como foi visto anteriormente, quando se utiliza um número elevado de quadros de referência na estimação de movimento, os desempenhos dos método “Entrelaçado”, “Entrelaçado” com a ferramenta PAFF e “Entrelaçado” com as ferramentas PAFF e MBAFF são similares.

As TAB. 4.1 e TAB. 4.2 apresentam alguns resultados obtidos com os método *Simulcast* (I), “Concatenado” (II) e “Entrelaçado” (III) nas simulações da seqüência MAN com os perfil *Main* e *High* usando o GoP II, para valores de PSNR e taxa fixas.

Na simulação da seqüência MAN no perfil *Main* com GoP II, o codificador teve dificuldades de manter ou alcançar algumas das taxas simuladas. Isto é justificado pelo fato de que o codificador com este GoP II apresenta uma variação lenta dos PQ utilizados na codificação dos quadros da seqüência e também pelo tamanho da mesma. Na codificação das vistas direitas o método “Entrelaçado” apresentou um desempenho similar para taxas de codificação superiores à 400 Kbps. Este desempenho similar ocorre devido a *flag IntraBotton* apresentar erros na codificação com taxa de bits constante.

Na simulação da seqüência MAN no perfil *High* com GoP I, todos os três métodos apresentaram um ganho de 0,2 dB. As curvas de desempenho dos métodos são similares as curvas do perfil *Main* com o GoP I acrescidas de 0,2 dB. Este ganho foi obtido devido a utilização da transformada adaptativa 4x4/8x8, pois a seqüência apresenta algumas regiões homogêneas (plano de fundo da imagem e algumas áreas do rosto). E utilizando este mesmo perfil com o GoP II, o método entrelaçado manteve o mesmo ganho de 0,2 dB nas seqüências esquerda e direita, mas os métodos “Concatenado” e *Simulcast*, nas seqüências esquerdas obtiveram ganhos aproximadamente de 0,6 e 0,8 dB para taxas de codificação baixas como 300 Kbps e 0,8 dB para taxas altas como 1000 Kbps. Os três métodos apresentaram resultados similares na codificação da vista direita.

TAB. 4.1: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP II. TAXA FIXA - MAN

MTD			I	II	III	Ganho (dB)		
Perfil	PQ	GoP	TAXA	TAXA	TAXA	III/II	III/I	II/I
-	-	-	E/D (Kbps)	E/D (Kbps)	E/D (Kbps)	E/D	E/D	E/D
<i>Main</i>	23	II	230.0	230.0	230.0	0.5/0.4	0.7/0.4	0.2/0.0
<i>Main</i>	23	II	1000.0	1000.0	1000.0	0.35/0.0	0.35/0.0	0.0/0.0
<i>High</i>	23	II	400.0	400.0	400.0	0.35/0.0	0.35/0.0	0.0/0.0
<i>High</i>	23	II	1000.0	1000.0	1000.0	0.30/0.0	0.30/0.0	0.0/0.0

TAB. 4.2: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP II. PSNR FIXO - MAN

MTD			I	II	III	Ganho (Kbps)		
Perfil	PQ	GoP	PSNR	PSNR	PSNR	III/II	III/I	II/I
-	-	-	E/D (dB)	E/D (dB)	E/D (dB)	E/D	E/D	E/D
<i>Main</i>	23	II	44.5	44.5	44.5	110.0/0.0	110.0/0.0	0.0/0.0
<i>High</i>	23	II	44.5	44.5	44.5	110.0/0.0	120.0/0.0	0.0/0.0

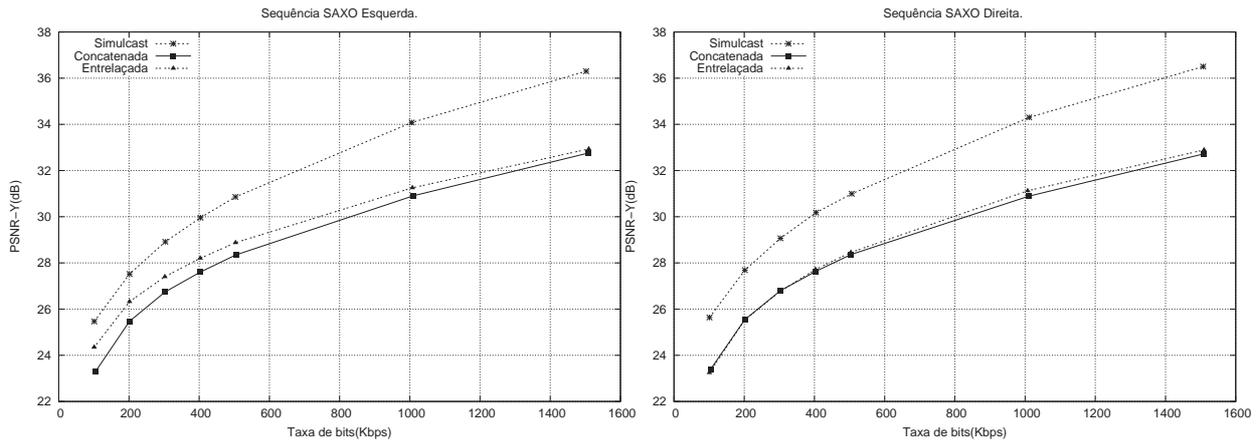
#### 4.4.2 RESULTADOS DA SEQUÊNCIA SAXO COM O PERFIL *MAIN* E *HIGH*.

Na codificação da sequência SAXO, o valor do parâmetro de quantização escolhido para a comparação entre os métodos foi o 39, por apresentar na média a menor variação da qualidade (PSNR-Y) entre os quadros codificados.

As FIG. 4.8a e FIG. 4.8b mostram as curvas de PSNR-Y x taxa de bits da sequência SAXO codificada através dos métodos *Simulcast*, “Concatenado” e “Entrelaçado”, sendo o último sem a ferramenta PAFF habilitada.

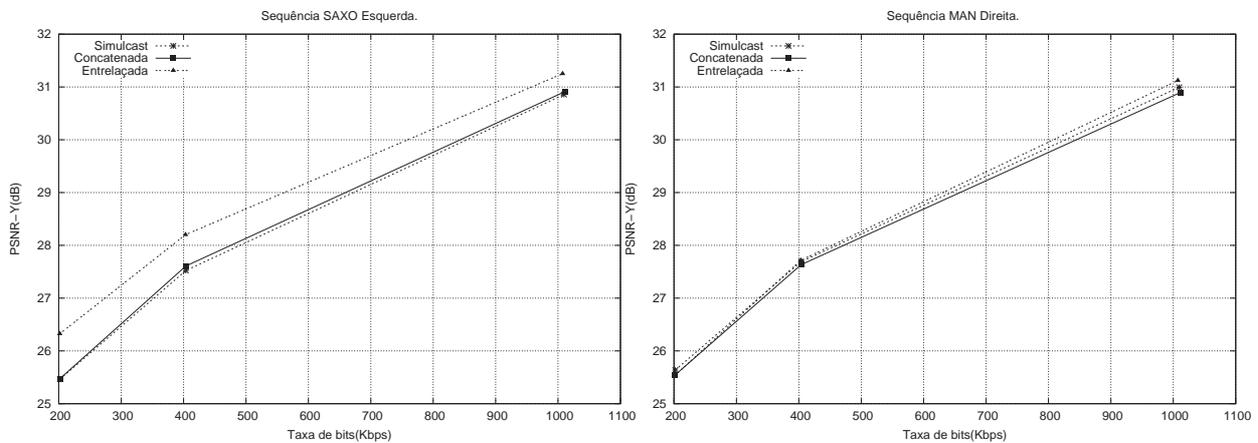
Podemos observar nas FIG. 4.8a e FIG. 4.8b que o método “Entrelaçado” apresenta um ganho para taxas de codificação baixas, como 100 Kbps, de aproximadamente 1,0 dB em relação ao método “Concatenado” nas sequências esquerdas e obteve um resultado similar nas sequências direitas. O método “Entrelaçado” apresenta um desempenho similar nos resultados das sequências direitas porque a *flag IntraBotton* apresenta erros quando o codificador é configurado para taxas de codificação constantes.

As FIG. 4.8c e FIG. 4.8d mostram as curvas de PSNR-Y x taxa de bits de 200, 400 e 1000 Kbps da sequência SAXO codificada através dos métodos *Simulcast*, “Concatenado”



(a) Sequência SAXO esquerda

(b) Sequência SAXO direita



(c) Sequência SAXO esquerda

(d) Sequência SAXO direita

FIG. 4.8: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” sem PAFF. Sequência SAXO - GoP I PQ39 - Perfil *Main*

e “Entrelaçado”, sendo o último sem a ferramenta PAFF habilitada. O eixo horizontal do gráfico é relativo à taxa de codificação da seqüência binocular (vistas esquerda e direita).

O método “Entrelaçado” apresenta o melhor desempenho. Na FIG. 4.8c vemos que os métodos *Simulcast* e “Concatenado” apresentam resultados similares e que o método “Entrelaçado”, em relação os dois anteriores, produz um ganho de aproximadamente 1,0 dB para uma taxa de codificação de 200 Kbps e 0,4 dB para 1000 Kbps. Para um valor de PSNR-Y igual 28,0 dB o ganho de codificação do método “Entrelaçado” em relação ao “Concatenado” e *Simulcast* é de aproximadamente 90 Kbps e 110 Kbps. Na FIG. 4.8d para uma taxa de 1000 Kbps, o método “Entrelaçado” obtém um ganho sobre os métodos *Simulcast* e “Concatenado” de 0,1 dB e 0,3 dB. O desempenho superior do

método “Entrelaçado” é dado pela possibilidade de utilizar na predição da vista direita (campo ímpar) a própria vista esquerda já codificada do quadro entrelaçado, como quadro de referência. Isto pode ser observado nos arquivos de dados estatísticos do processo de codificação gerados pelo codificador, mostrando que o método “Entrelaçado” por ter um maior número de quadros de referência realiza um maior número de predição Inter-quadros do que os métodos “Concatenado” e *Simulcast*. Quando a taxa de codificação é aumentada, o desempenho dos métodos “Concatenado” e *Simulcast* melhora, pois os mesmos passam a realizar mais predições inter-quadros com tamanhos de blocos 8x8, aumentando a acurácia do processo de estimação de movimento.

Foram testadas mais duas configurações com a seqüência SAXO. A primeira com o perfil *Main* com GoP II e a segunda no perfil *High* com GoP I. As curvas dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” nestas simulações apresentaram um comportamento similar ao descrito para as curvas com o perfil *Main* com GoP I. As TAB. 4.3 e TAB. 4.4 apresenta alguns resultados obtidos com os métodos *Simulcast*, “Concatenado” e “Entrelaçado” nas simulação da seqüência SAXO com perfil *Main* usando o GoP II.

Todos os três métodos com o perfil *High* apresentaram um ganho máximo de aproximadamente 0,3 dB. As observações feitas para os métodos no perfil *Main* são as mesmas para o perfil *High*.

TAB. 4.3: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main*. GoP II. TAXA FIXA - SAXO

MTD			I	II	III	Ganho (dB)		
Perfil	PQ	GoP	TAXA	TAXA	TAXA	III/II	III/I	II/I
-	-	-	E/D (Kbps)	E/D (Kbps)	E/D (Kbps)	E/D	E/D	E/D
<i>Main</i>	39	II	210.0	210.0	210.0	0.8/0.5	0.8/0.5	0.0/0.0
<i>Main</i>	39	II	500.0	500.0	500.0	0.6/0.25	0.6/0.25	0.0/0.0
<i>Main</i>	39	II	1000.0	1000.0	1000.0	0.65/0.35	0.65/0.35	0.0/0.0

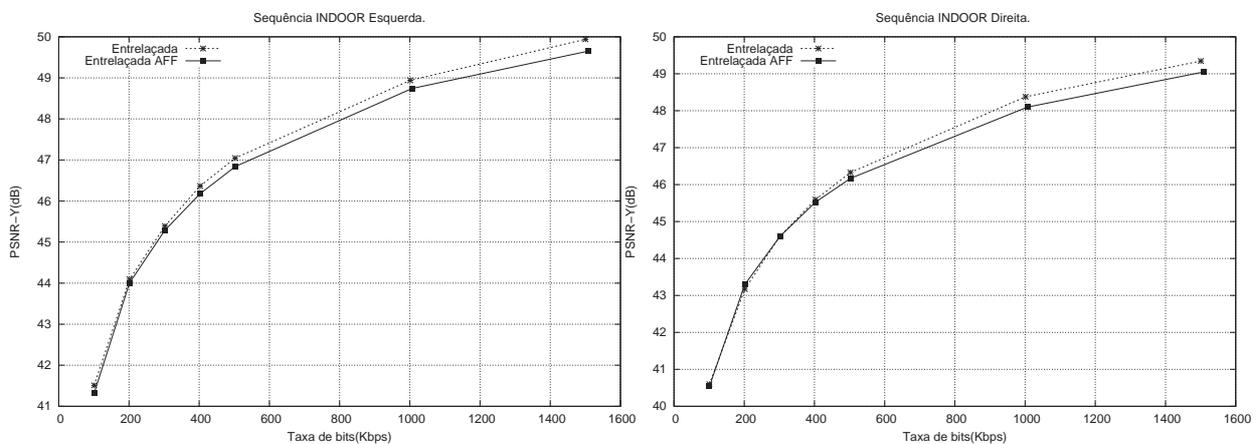
#### 4.4.3 RESULTADOS DA SEQÜÊNCIA INDOOR COM O PERFIL *MAIN* E *HIGH*.

O valor do parâmetro de quantização escolhido para a comparação entre os métodos foi o 23 por apresentar na média a menor variação da qualidade (PSNR-Y) entre os quadros codificados.

TAB. 4.4: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main*. GoP II. PSNR FIXO - SAXO

MTD			I	II	III	Ganho (Kbps)		
Perfil	PQ	GoP	PSNR	PSNR	PSNR	III/II	III/I	II/I
-	-	-	E/D (dB)	E/D (dB)	E/D (dB)	E/D	E/D	E/D
<i>Main</i>	39	II	28.0	28.0	28.0	80.0/50.0	80.0/50.0	0.0/0.0

As FIG. 4.9a e FIG. 4.9b mostram as curvas de PSNR-Y x taxa de bits da seqüência INDOOR codificada através do método “Entrelaçado” com e sem a ferramenta PAFF habilitada.



(a) Seqüência INDOOR esquerda.

(b) Seqüência INDOOR direita.

FIG. 4.9: Comparação do Método “Entrelaçado” com/sem PAFF. GoP I PQ23 - Perfil *Main* - INDOOR

Podemos observar que o método “Entrelaçado sem a ferramenta PAFF apresenta um resultado superior ou similar nas seqüência esquerda e direita. A explicação para este resultado é a mesma descrita para as seqüências MAN e SAXO. Como no processo de codificação pode-se utilizar um número grande de quadros de referência, o desempenho do método “Entrelaçado” com ou sem a ferramenta PAFF é similar. A diferença das curvas é devido a tentativa do codificador em manter a taxa de bits constante. O método “Entrelaçado” sem a ferramenta PAFF apresenta uma variação mais suave de qualidade entre quadros adjacentes no processo de codificação.

Temos nas FIG. 4.10a e FIG. 4.10b as curvas de PSNR-Y x taxa de bits da seqüência INDOOR codificada através dos método *Simulcast*, “Concatenado” e “Entrelaçado”,

sendo o último sem a ferramenta PAFF.

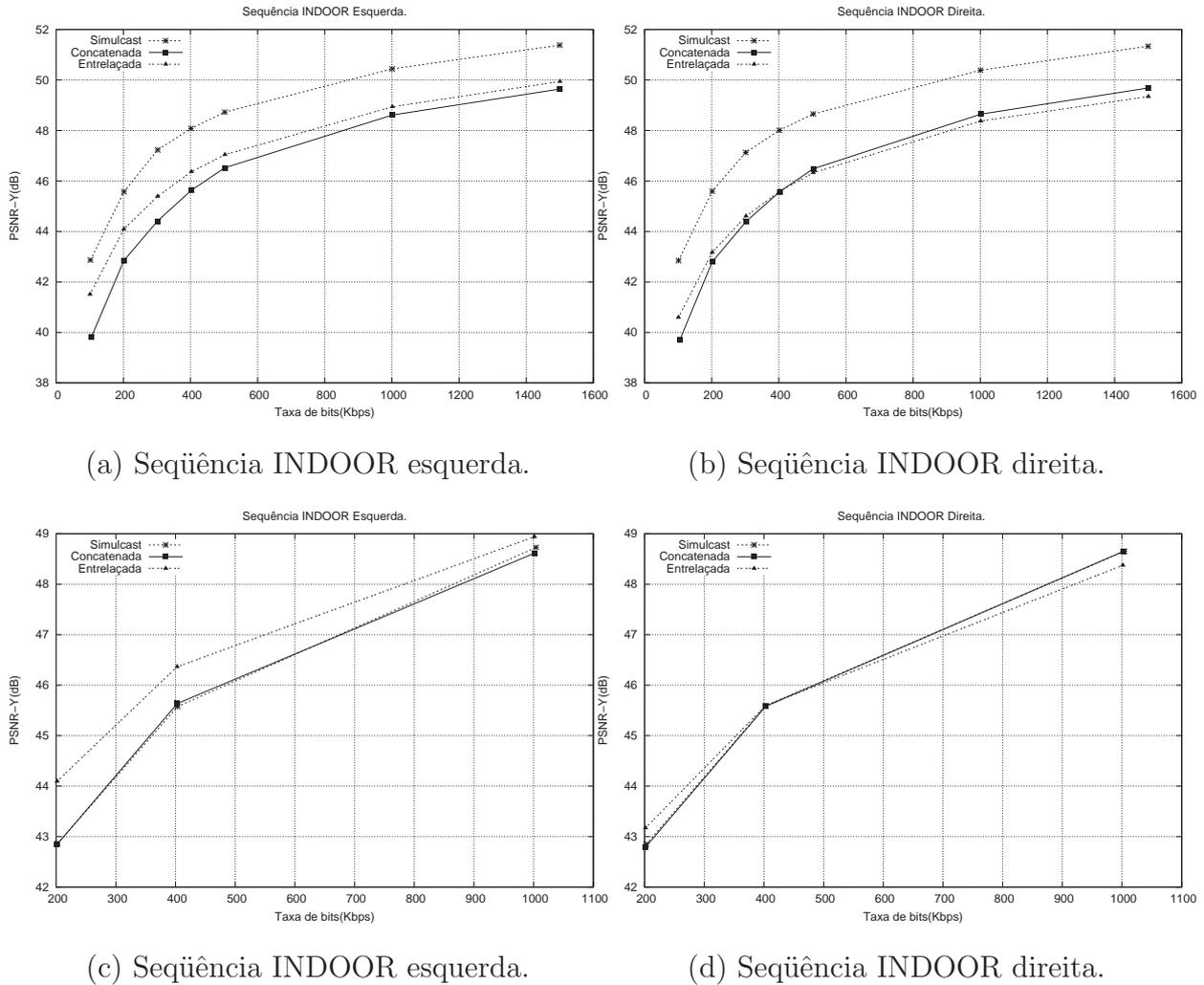


FIG. 4.10: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” sem PAFF. Sequência INDOOR - GoP I PQ23 - Perfil *Main*.

Podemos observar nas FIG. 4.10a que o método “Entrelaçado” apresenta um ganho para taxas de codificação baixas como 110 Kbps de aproximadamente 1,75 dB em relação ao método “Concatenado” nas seqüências esquerdas, e de 1,0 dB nas seqüências direitas e para taxas altas como 1500 Kbps nas seqüências esquerdas de 0,2 dB. Na seqüência direita (FIG. 4.10b) o método “Entrelaçado” apresenta um resultado superior ao método ‘Concatenado’ somente para taxas de codificação inferiores a 400 Kbps. As FIG. 4.10c e FIG. 4.10d mostram as curvas de PSNR-Y x taxa de bits de 200, 400 e 1000 Kbps da seqüência INDOOR codificada através dos método *Simulcast*, “Concatenado” e “Entrelaçado”, sendo o último sem a ferramenta PAFF. O eixo horizontal do gráfico é relativo

à taxa de codificação da seqüência binocular (vistas esquerda e direita).

O método “Entrelaçado” apresenta o melhor desempenho para a seqüência esquerda. Na FIG. 4.10c vemos que os métodos *Simulcast* e “Concatenado” apresentam resultados similares, e que, em relação os dois métodos anteriores o método “Entrelaçado” obtém um ganho de aproximadamente 1,2 dB para uma taxa de codificação de 200 Kbps, e 0,25 dB (*Simulcast*) e 0,3 dB (“Concatenado”) para 1000 Kbps. E na FIG. 4.10d o método “Entrelaçado” obtém um ganho sobre os métodos *Simulcast* e “Concatenado” para uma taxa de 200 Kbps de 0,4 dB, sendo que para uma taxa acima de 400 Kbps os métodos *Simulcast* e “Concatenado” apresentam um desempenho superior de até 0,4 dB para uma taxa de 1000 Kbps. Para um valor de PSNR-Y igual a 46,0 dB, o método “Entrelaçado” apresenta um ganho de taxa de codificação de aproximadamente 120 Kbps em relação aos métodos “Concatenado” e *Simulcast* nas seqüências esquerdas. A seqüência estéreo INDOOR apresenta valores de disparidades baixos. O desempenho superior do método “Entrelaçado” nas seqüências esquerdas é justificado pelos autores em (SUN, 2003) e pela análise realizada nesta dissertação, aonde vimos que o ganho do método “Entrelaçado” é obtido principalmente da possibilidade da vista direita (campo ímpar) utilizar a vista esquerda codificada (campo par) como quadro de referência na predição entre campos, para seqüências estéreo com valores de disparidade baixo. O desempenho inferior do método “Entrelaçado” nas seqüências direitas é devido a *flag IntraBotton* não poder ser utilizada na codificação com taxa constante. Analisando os arquivos de dados estatísticos do processo de codificação gerados pelo codificador dos métodos *Simulcast* e “Concatenado”, podemos observar que a média de bits por quadro utilizada pelo método “Concatenado” é praticamente a soma das médias das seqüências esquerdas e direitas utilizando o método *Simulcast*.

Na codificação da seqüência INDOOR utilizando o GoP II com um PQ de valor 23, o codificador não manteve à taxa de codificação constante, diferente das simulações com o GoP I, onde esse problema não foi observado. Houve a necessidade, neste caso de aumentar o valor do PQ. O valor escolhido foi o de 30, por proporcionar a menor oscilação do valor da taxa de codificação pré-determinada para o GoP II.

As TAB. 4.5 e TAB. 4.6 apresentam alguns resultados obtidos com os métodos *Simulcast*, “Concatenado” e “Entrelaçado” nas simulações da seqüência INDOOR no perfil *Main* com GoP II e também no perfil *High* com GoP I.

Na simulação do método “Entrelaçado” com o perfil *High* com GoP I, o mesmo apre-

sentou ganhos de 3,2 dB e 2,2 dB para 100 e 1500 Kbps nas seqüências esquerdas e 3,6 dB e 2,3 dB nas vistas direitas para as mesmas taxas de codificação em relação a simulação utilizando o perfil *Main*. Os ganhos presentes nos métodos “Concatenado” e “Entrelaçado” foram no máximo de 0,2 dB. Na codificação do método “Entrelaçado”, o codificador apresentou uma variação suave nos PQ dos quadros adjacentes, resultando em um ganho de desempenho pelo método.

TAB. 4.5: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP I e II. TAXA FIXA - INDOOR

MTD			I	II	III	Ganho (dB)		
Perfil	PQ	GoP	TAXA	TAXA	TAXA	III/II	III/I	II/I
-	-	-	E/D (Kbps)	E/D (Kbps)	E/D (Kbps)	E/D	E/D	E/D
<i>Main</i>	23	II	300.0	300.0	300.0	0.5/0.45	0.5/0.45	0.0/0.0
<i>Main</i>	23	II	500.0	500.0	500.0	0.37/0.37	0.3/0.37	0.07/0.0
<i>Main</i>	23	II	1000.0	1000.0	1000.0	0.0/0.2	0.0/0.2	0.0/0.0
<i>High</i>	23	I	300.0	300.0	300.0	3.6/3.7	3.6/3.7	0.0/0.0
<i>High</i>	23	I	500.0	500.0	500.0	3.2/3.3	3.2/3.3	0.0/0.0
<i>High</i>	23	I	1000.0	1000.0	1000.0	3.6/3.6	3.6/3.6	0.0/0.0

TAB. 4.6: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP I e II. PSNR FIXO - INDOOR

MTD			I	II	III	Ganho (Kbps)		
Perfil	PQ	GoP	PSNR	PSNR	PSNR	III/II	III/I	II/I
-	-	-	E/D (dB)	E/D (dB)	E/D (dB)	E/D	E/D	E/D
<i>Main</i>	23	II	46.00	46.00	46.00	65.0/65.0	50.0/65.0	5.0/0.0
<i>High</i>	23	I	47.0	47.0	47.0	110.0/0.0	120/0.0	0.0/0.0

#### 4.4.4 RESULTADOS DA SEQÜÊNCIA OUTDOOR COM O PERFIL *MAIN* E *HIGH*.

As FIG. 4.11a e FIG. 4.11b mostram as curvas de PSNR-Y x taxa de bits da seqüência OUTDOOR codificada através do método “Entrelaçado” com a ferramenta PAFF habilitada. Podemos observar que o método “Entrelaçado sem a ferramenta PAFF apresenta

um resultado superior ou similar nas seqüências esquerdas e direitas. Como já comentado anteriormente, quando o método “Entrelaçado” é configurado com a possibilidade de utilizar um número grande de quadros de referência para a estimação de movimento, o desempenho dos métodos “Entrelaçado”, “Entrelaçado” com a ferramenta PAFF habilitada e com as ferramentas PAFF e MBAFF habilitadas, obtêm desempenhos similares. A justificativa para este resultado está relacionada com a variação do PQ para manter a taxa constante pré-definida. No processo de codificação do método “Entrelaçado” sem a ferramenta PAFF habilitada, há uma variação suave no valor do parâmetro de quantização na codificação de quadros adjacentes. Com isso, há um aumento na qualidade dos quadros codificados.

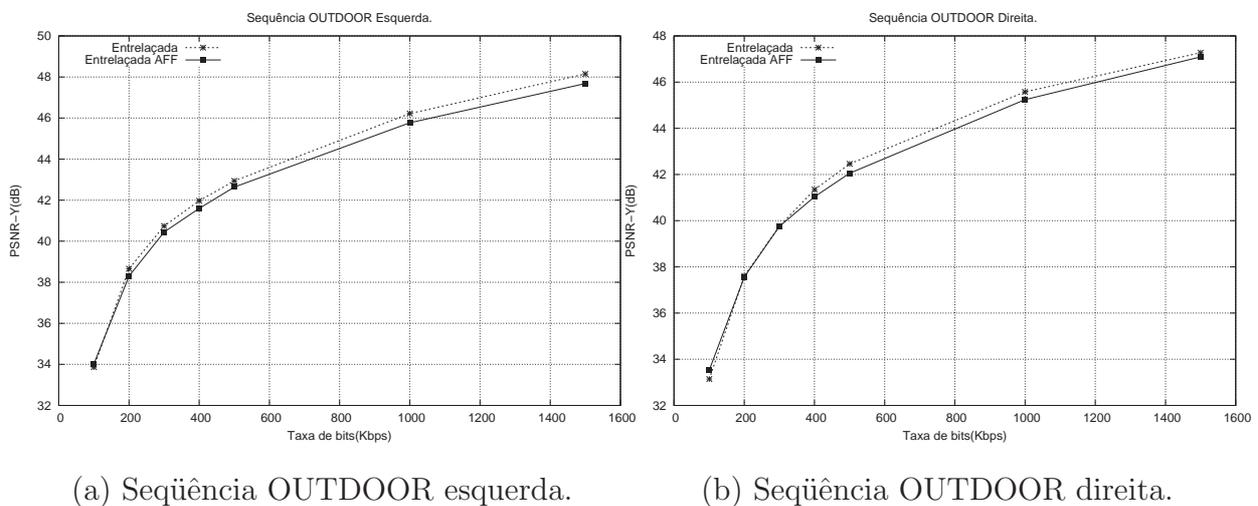
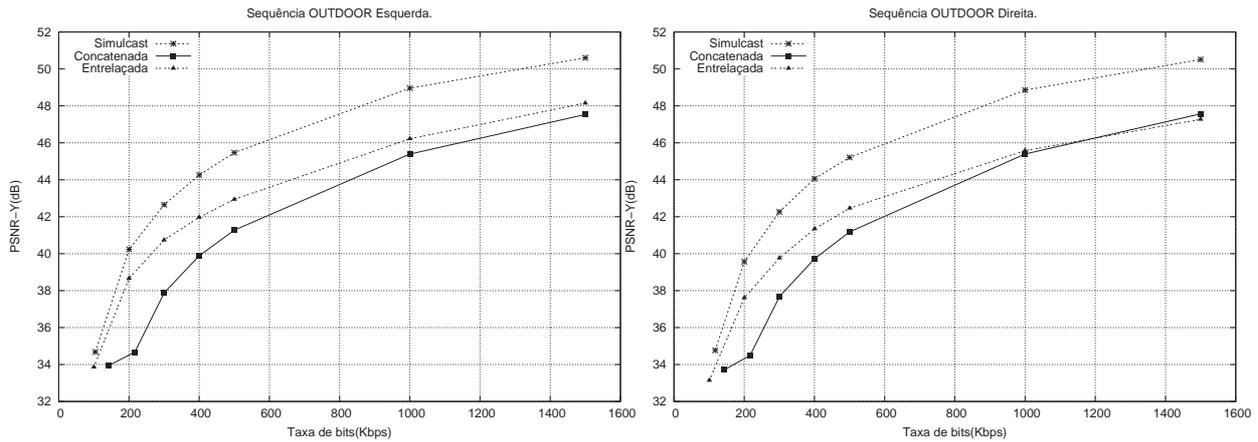


FIG. 4.11: Comparação do Método “Entrelaçado” com/sem PAFF. GoP I PQ23 - Perfil *Main* - OUTDOOR

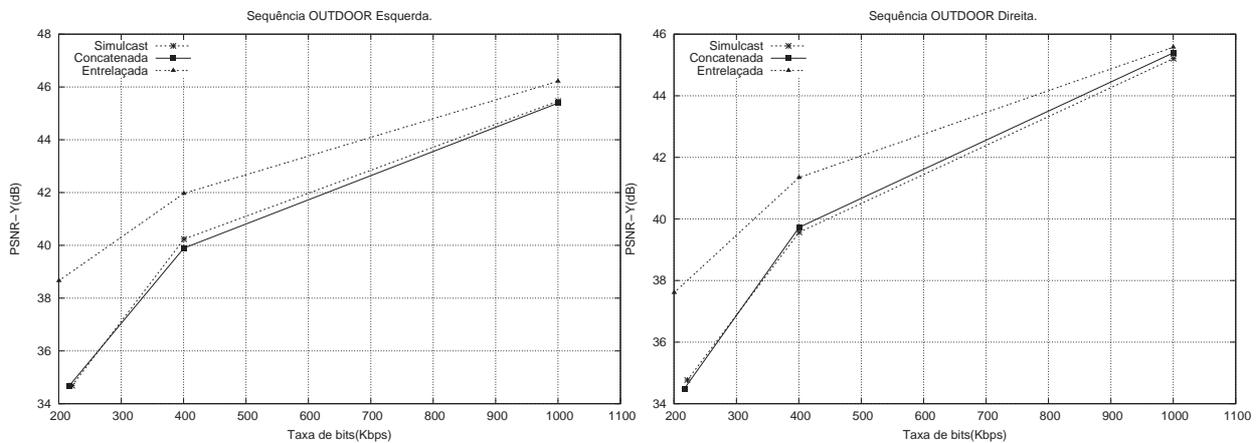
As FIG. 4.12a e FIG. 4.12b mostram as curvas de PSNR-Y x taxa de bits da seqüência OUTDOOR codificada através dos método *Simulcast*, “Concatenado” e “Entrelaçado”, sendo o último sem a ferramenta PAFF.

Podemos observar nas FIG. 4.12a e FIG. 4.12b que, para taxas de codificação baixas, como 100 e 200 Kbps, o codificador apresenta dificuldade para manter à taxa de codificação constante. O método “Entrelaçado” apresenta um ganho para taxas de codificação baixas como 200 e 400 Kbps de aproximadamente 4,0 e 2,0 dB em relação ao método “Concatenado” na seqüência esquerda. Para taxas altas, como 1500 Kbps, o ganho foi de 0,6 dB. Na seqüência direita obteve um resultado de 3,2 e 1,8 dB para taxas de codificação baixas como 200 e 400 Kbps. Para taxas altas acima de 1200 Kbps o desempenho do método



(a) Sequência OUTDOOR esquerda.

(b) Sequência OUTDOOR direita.



(c) Sequência OUTDOOR esquerda.

(d) Sequência OUTDOOR direita.

FIG. 4.12: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” sem PAFF. Sequência OUTDOOR - GoP I PQ23 - Perfil *Main*.

Entrelaçado foi pior ao que a do método “Concatenado”. Como não foi possível habilitar a *IntraBotton*, o desempenho das seqüências direitas codificadas através do método “Entrelaçado” é inferior aos das seqüências esquerdas. Podemos observar também que com o aumento da taxa de codificação, os desempenhos dos métodos “Concatenado” e *Simulcast* aumentam, pois analisando os arquivos de dados estatísticos do processo de codificação gerado pelo codificador, vemos que os mesmo passam a realizar mais predições inter-quadros com blocos de tamanhos pequenos. Desta forma, aumentando a qualidade dos quadros em regiões de movimento. As FIG. 4.12c e FIG. 4.12d mostram as curvas de PSNR-Y x taxa de bits de 200, 400 e 1000 Kbps da seqüência OUTDOOR codificada através dos método *Simulcast*, “Concatenado” e “Entrelaçado”, sendo o último sem a fer-

ramenta PAFF. O eixo horizontal do gráfico é relativo à taxa de codificação da seqüência binocular (vistas esquerda e direita).

Na FIG. 4.12c vemos que o método *Simulcast* apresenta resultados similares ou superiores ao método “Concatenado” e que o método “Entrelaçado” em relação os dois anteriores obtém um ganho de aproximadamente 3,1 dB, 1,8 dB e 0,8 dB para uma taxa de codificação de 300, 400 e 1000 Kbps. Na FIG. 4.12d o método “Entrelaçado” apresenta um ganho sobre os métodos *Simulcast* e “Concatenado” para taxas de 300, 400 e 1000 Kbps de 2,5 dB, 1,8 dB e 0,3 dB (“Concatenado”) e 0,25 dB (“Entrelaçado”). O método *Simulcast* apresenta um resultado similar ou inferior ao “Concatenado” nas taxas codificadas. Para um valor de PSNR-Y igual a 42,0 dB, o método “Entrelaçado” obteve um ganho de codificação de aproximadamente 200 Kbps na seqüência esquerda, em relação aos dois métodos comparados, e de 40 Kbps (“Concatenado”) e 60 Kbps (*Simulcast*) na seqüência direita.

A seqüência OUTDOOR também foi simulada no perfil *High* com GoP I. O comportamento das curvas dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” é similar ao descrito para as curvas com o perfil *Main* com GoP I. As TAB. 4.7 e TAB. 4.8 apresentam alguns resultados obtidos com os métodos *Simulcast*, “Concatenado” e “Entrelaçado” nas simulações.

TAB. 4.7: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *High*. GoP I. TAXA FIXA - OUTDOOR

MTD			I	II	III	Ganho (dB)		
Perfil	PQ	GoP	TAXA	TAXA	TAXA	III/II	III/I	II/I
-	-	-	E/D (Kbps)	E/D (Kbps)	E/D (Kbps)	E/D	E/D	E/D
<i>High</i>	23	I	300.0	300.0	300.0	3.0/2.5	3.0/0.2	0.0/-2.3
<i>High</i>	23	I	1000.0	1000.0	1000.0	1.8/1.0	1.8/0.2	0.0/-1.6

TAB. 4.8: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *High*. GoP I. PSNR FIXO - OUTDOOR

MTD			I	II	III	Ganho (Kbps)		
Perfil	PQ	GoP	PSNR	PSNR	PSNR	III/II	III/I	II/I
-	-	-	E/D (dB)	E/D (dB)	E/D (dB)	E/D	E/D	E/D
<i>High</i>	23	I	42.0	42.0	42.0	280.0/300.0	260.0/330.0	-20.0/30.0

Os resultados obtidos para todas as seqüências simuladas indicam que o desempenho do método “Entrelaçado” será mantido para outras taxas de codificação.

#### 4.5 COMPARAÇÃO ENTRE OS MÉTODOS - TAXA DE BITS VARIÁVEL

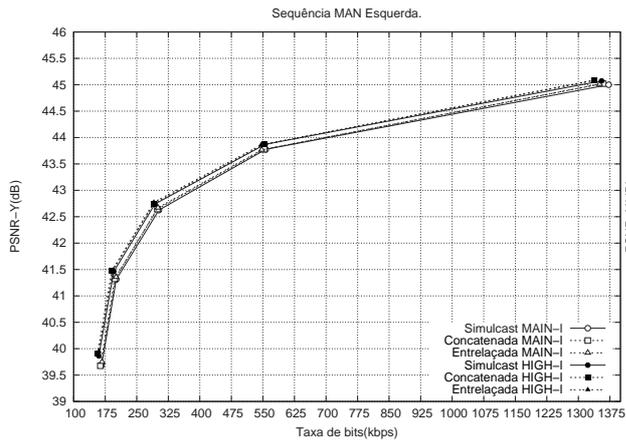
Os métodos *Simulcast*, “Concatenado” e “Entrelaçado” foram simulados nos perfis *Main* e *High*, sendo no último configurada a opção de transformada adaptativa de blocos 4x4/8x8. No método “Entrelaçado” a opção de codificação por campo foi configurada. O nível utilizado foi o mesmo das simulações com taxa de bits constante. Os GoPs II (IBBPBBP...) e III (IBPBPBP...), com os parâmetros de quantização (PQ) {16, 21, 26, 31, 36, 41} para MAN, INDOOR e OUTDOOR, e {21, 26, 31, 36, 41, 46} para SAXO, foram utilizados em todas as simulações.

A *flag IntraBottom* foi habilitada. Agora no método “Entrelaçado”, na codificação de um quadro tipo I ambos os campos são codificação como Intra.

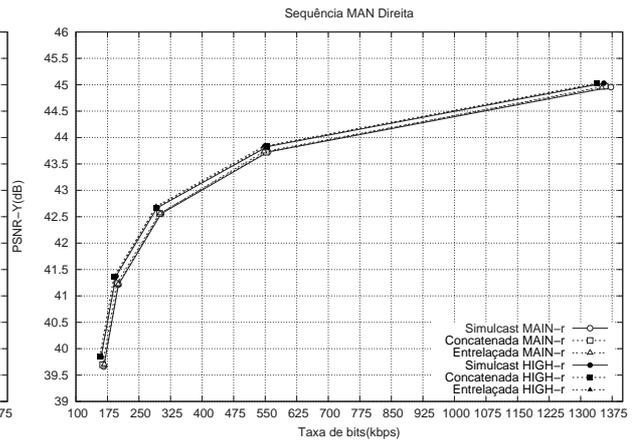
O desempenho do método “Entrelaçado” é similar aos métodos “Concatenado” e *Simulcast* devido o fato de que nos parâmetros de configuração, o número de quadros de referência utilizados para a codificação de quadros tipo B (predição bi-lateral) foi configurado capacitando o codificador a utilizar cinco quadros de referência anteriores e apenas um para quadro posterior. Com isso, na codificação da seqüências MAN utilizando o método “Entrelaçado” há uma perda de desempenho na codificação dos quadros tipo B, pois com a possibilidade de usar apenas um quadro posterior como referência, a vista esquerda usa a vista direita codificada do mesmo quadro desmontado, e nem sempre há um bom casamento na estimação de movimento devido ao grau de distorção perspectiva presente no par estéreo. Outro fator está também na codificação da vista direita que utiliza a vista esquerda codificada do quadro posterior, pois além da distorção perspectiva, há um deslocamento temporal da imagem.

##### 4.5.1 RESULTADOS - SEQÜÊNCIA MAN.

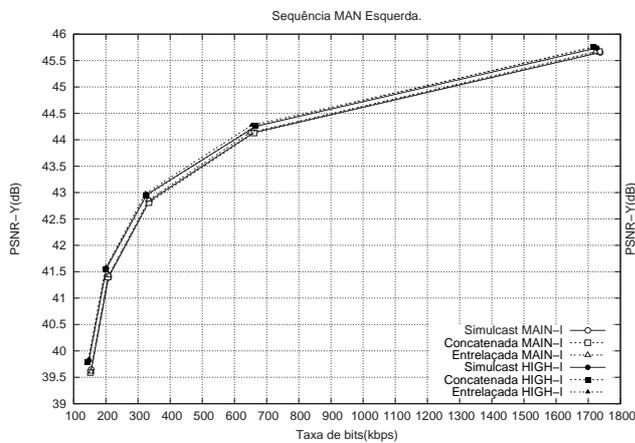
A FIG. 4.13 apresenta os resultados das simulações dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” para os GoPs II e III nos perfis *Main* e *High*. O eixo horizontal do gráfico (Taxa de Bits) representa à taxa de codificação total da seqüência binocular.



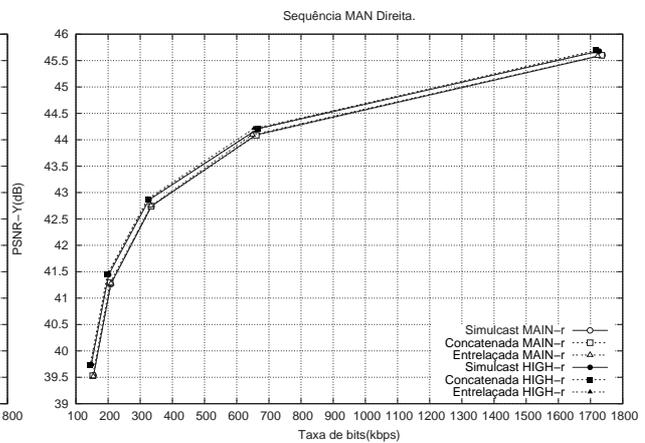
(a) Sequência MAN esquerda - GoP II.



(b) Sequência MAN direita - GoP II.



(c) Sequência MAN esquerda - GoP III.



(d) Sequência MAN direita - GoP III.

FIG. 4.13: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil *MainxHigh* - MAN.

Pode-se observar que os resultados obtidos para os métodos nos dois perfis apresentaram resultados similares. O pequeno ganho presente nas curvas simuladas no perfil *High* é devido ao fato de que neste perfil a transformada adaptativa 4x4/8x8 foi habilitada.

Os valores da taxa de bits utilizada nos métodos *Simulcast* (taxa de bits da vista esquerda somada à taxa de bits da vista direita) e “Concatenado” (taxa de bits da seqüência “concatenada”) no perfil *Main* e *High* com estrutura de GoP II e III foram similares. Como foi visto nos resultados obtidos na codificação com taxa de bits constante, os métodos “Concatenado” e *Simulcast* apresentam desempenho similar. As simulações utilizando o GoP III apresentaram resultados superiores na qualidade das seqüências reconstruídas em relação as simulações realizadas com o GoP II, devido o GoP III apresentar um estrutura

que apresenta um número menor de quadros tipo B (PQ constante = 30) e de tamanho menor (número maior de quadros tipo I).

O método “Entrelaçado” utiliza na média uma taxa de bits inferior aos demais e seu desempenho é superior aos métodos *Simulcast* e “Concatenado” (TAB. 4.9 - FIG. 4.13).

As TAB. 4.9 e TAB. 4.10 apresentam alguns resultados selecionados obtidos para os três métodos, com valores similares de qualidade das seqüências reconstruídas (PSNR-Y), dos GoP II e III, para a comparação da taxa de compressão e do fator de uso de banda. Nas TAB. 4.9 e TAB. 4.10, os métodos (MTD) *Simulcast*, “Concatenado” e “Entrelaçado” são representados pelos algarismo romanos I, II e III.

TAB. 4.9: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP II. - MAN

MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	16	45.00	44.96	45.02	44.97	45.01	44.94	80.29	79.71	81.18	90.53	1.015	1.132
<i>High</i>	16	45.07	45.02	45.09	45.03	45.08	45.01	81.48	80.88	82.40	82.40	1.015	1.015
<i>Main</i>	21	43.78	43.73	43.78	43.72	43.77	43.73	200.00	196.43	200.00	201.83	1.009	1.018
<i>High</i>	21	43.87	43.83	43.87	43.84	43.86	43.83	200.00	196.43	198.20	201.83	1.000	1.018
<i>Main</i>	26	42.64	42.56	42.65	42.56	42.67	42.56	366.67	354.84	366.67	366.67	1.017	1.017
<i>High</i>	26	42.74	42.67	42.75	42.67	42.77	42.70	366.67	366.67	379.31	379.31	1.035	1.035
<i>Main</i>	31	41.31	41.21	41.32	41.24	41.36	41.22	523.81	550.00	550.00	550.00	1.025	1.025
<i>High</i>	31	41.46	41.37	41.47	41.37	41.49	41.37	550.00	550.00	578.95	564.10	1.053	1.026
<i>Main</i>	36	39.69	39.66	39.68	39.70	39.75	39.70	647.06	647.06	666.67	647.06	1.030	1.000
<i>High</i>	36	39.86	39.85	39.90	39.85	39.95	39.83	687.50	687.50	687.50	687.50	1.000	1.000

Podemos observar nas TAB. 4.9 e TAB. 4.10 para valores de PSNR-Y similares, que o método “Entrelaçado” (III) apresenta na média a maior taxa de codificação (TC), tendo o método “Concatenado” (II) o segundo melhor desempenho. E analisando os resultados obtidos em relação ao fator de uso de banda, verificamos que os métodos “Entrelaçado” e “Concatenado” geram trens de bits menores em relação ao método *Simulcast* (I).

#### 4.5.2 RESULTADOS - SEQÜÊNCIA SAXO.

A FIG. 4.14 apresenta os resultados das simulações dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” para os GoP II e III nos perfis *Main* e *High*. O eixo horizontal do

TAB. 4.10: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP III. - MAN

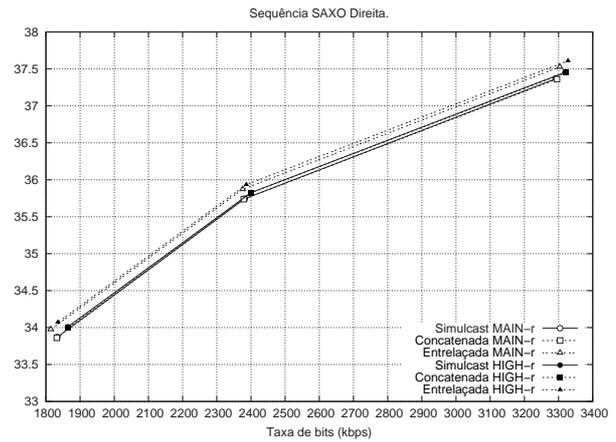
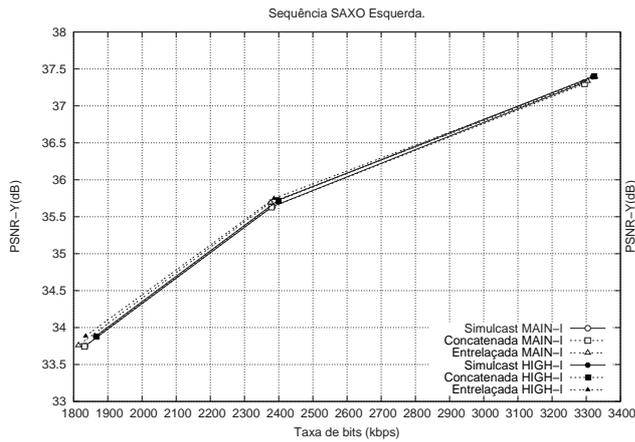
MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	16	45.65	44.60	45.67	45.60	45.67	45.59	61.11	60.77	61.11	61.62	1.003	1.011
<i>High</i>	16	45.74	45.68	45.69	45.76	45.74	45.67	61.45	60.77	61.80	61.80	1.011	1.011
<i>Main</i>	21	44.14	44.09	44.13	44.09	44.14	44.10	161.76	157.14	160.58	162.96	1.007	1.002
<i>High</i>	21	44.26	44.21	44.21	44.26	44.27	44.21	159.42	157.14	159.42	161.76	1.001	1.022
<i>Main</i>	26	42.84	42.74	42.81	42.74	42.86	42.75	314.29	314.29	314.29	318.84	1.000	1.022
<i>High</i>	26	42.96	42.86	42.86	42.94	42.98	42.88	323.53	323.53	323.53	328.36	1.000	1.019
<i>Main</i>	31	41.39	41.26	41.40	41.29	41.43	41.28	500.00	500.00	511.63	500.00	1.025	1.025
<i>High</i>	31	41.56	41.44	41.45	41.54	41.59	41.47	523.81	523.81	523.81	523.81	1.000	1.000
<i>Main</i>	36	39.65	39.53	39.59	39.53	39.61	39.52	647.06	687.50	687.50	687.50	1.031	1.013
<i>High</i>	36	39.82	39.74	39.73	39.79	39.84	39.74	687.50	687.50	733.33	709.68	1.067	1.032

gráfico (Taxa de Bits) representa à taxa de codificação total da seqüência binocular.

Os resultados obtidos para os métodos nos dois perfis apresentaram resultados similares. O pequeno ganho presente nas curvas simuladas no perfil *High* é devido ao fato de que neste perfil há a transformada adaptativa 4x4/8x8 foi habilitada. Com isso, os valores de qualidade das seqüências reconstruídas (PSNR-Y) para todas as simulações, foram similares. Os valores da taxa de bits utilizada nos métodos *Simulcast* (taxa de bits da vista esquerda somada à taxa de bits da vista direita) e “Concatenado” (taxa de bits da seqüência “concatenada”) no perfil *Main* e *High* com estrutura de GoP II e III foram similares. As simulações utilizando o GoP III apresentaram resultados superiores na qualidade das seqüências reconstruídas em relação as realizadas com o GoP II, devido o GoP III apresentar um estrutura que apresenta um número menor de quadros tipo B (PQ constante = 30) e de tamanho menor (número maior de quadros tipo I).

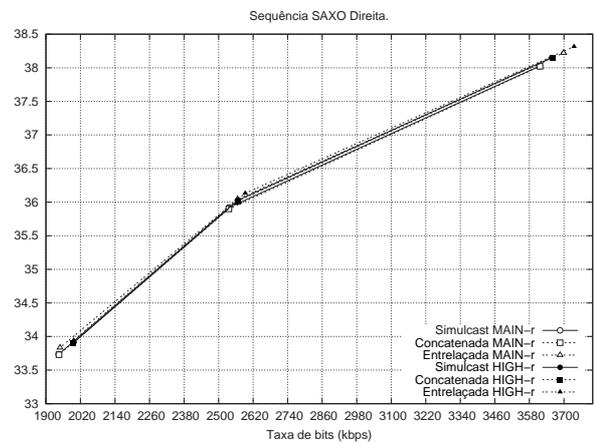
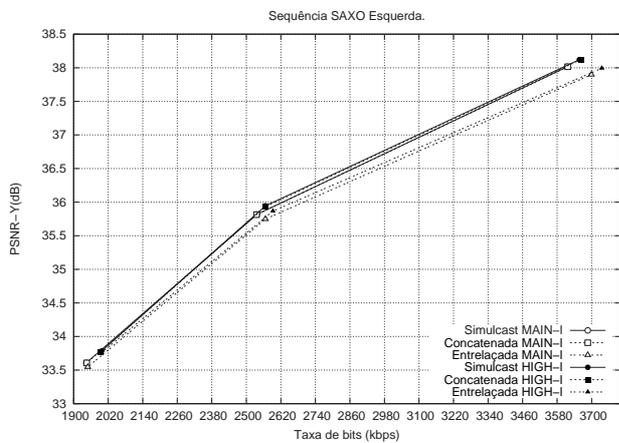
O método “Entrelaçado” (III) com GoP II é na média utiliza uma taxa de bits inferior aos demais e seu desempenho é superior aos métodos *Simulcast* (I) e “Concatenado” (II) (TAB. 4.9 - FIG. 4.14) e utilizando o GoP III, o método “Concatenado” na média utiliza uma taxa de bits inferior aos demais e seu desempenho é superior aos métodos.

As TAB. 4.11 e TAB. 4.12 apresentam alguns resultados selecionados obtidos para os três métodos, com valores similares de qualidade das seqüências reconstruídas (PSNR-Y), dos GoP II e III, para a comparação da taxa de compressão e do fator de uso de



(a) Sequência SAXO esquerda - GoP II.

(b) Sequência SAXO direita - GoP II.



(c) Sequência SAXO esquerda - GoP III.

(d) Sequência SAXO direita - GoP III.

FIG. 4.14: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil *MainxHigh* - SAXO.

banda. Nas TAB. 4.11 e TAB. 4.12, os métodos (MTD) *Simulcast*, “Concatenado” e “Entrelaçado” são representados pelos algarismo romanos I, II e III.

Na TAB. 4.11 para valores de PSNR-Y similares, que o método “Entrelaçado” (III) apresenta na média a maior taxa de codificação (TC), tendo o método “Concatenado” (II) o segundo melhor desempenho. E analisando os resultados obtidos em relação ao fator de uso de banda, verificamos que os métodos “Entrelaçado” e ”Concatenado” geram na média, trens de bits menores em relação ao método *Simulcast* (I). Utilizando o GoP III, o método “Concatenado” obtém na média o melhor resultados (TAB. 4.12) em relação aos demais métodos em relação a taxa de compressão e a utilização da banda.

TAB. 4.11: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP II. - SAXO

MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	21	37.31	37.37	37.30	37.36	37.34	37.53	22.22	22.84	22.83	22.76	1.000	0.997
<i>High</i>	21	37.39	37.45	37.40	37.45	37.39	37.61	22.02	22.63	22.62	22.59	1.000	0.999
<i>Main</i>	26	35.64	35.74	35.63	35.74	35.69	35.87	30.71	31.62	31.58	31.65	1.000	1.002
<i>High</i>	26	35.73	35.83	35.72	35.82	35.75	35.93	30.45	31.22	31.32	31.51	1.002	1.008
<i>Main</i>	31	33.75	33.87	33.75	33.86	33.76	33.98	39.78	40.88	40.98	41.44	1.003	1.014
<i>High</i>	31	33.88	34.01	33.88	34.00	33.88	34.07	39.15	40.22	40.32	40.98	1.003	1.019

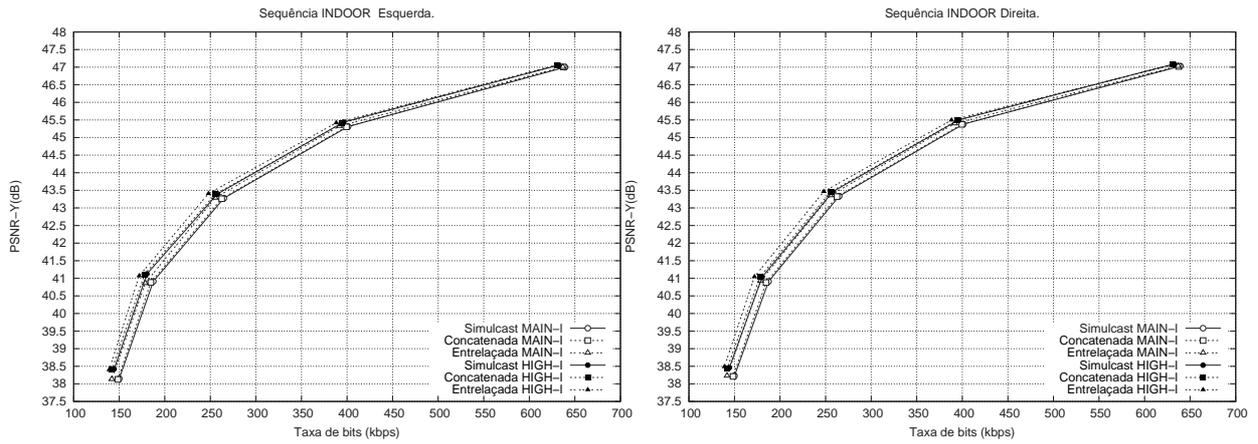
TAB. 4.12: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP III. - SAXO

MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	21	38.02	38.03	38.02	38.02	37.90	38.22	19.47	19.95	19.97	19.53	1.000	0.977
<i>High</i>	21	38.12	38.15	38.12	38.15	37.99	38.32	19.22	19.73	19.74	19.35	1.000	0.981
<i>Main</i>	26	35.82	35.92	35.82	35.90	35.75	36.04	27.82	28.35	28.52	28.14	1.002	0.989
<i>High</i>	26	35.95	36.03	35.93	36.01	35.87	36.13	27.41	28.14	28.14	27.88	1.000	0.991
<i>Main</i>	31	33.61	33.74	33.61	33.73	33.55	33.83	36.27	37.00	37.13	37.04	1.000	1.996
<i>High</i>	31	33.79	33.91	33.77	33.91	33.76	33.94	35.41	35.92	36.23	36.14	1.003	1.000

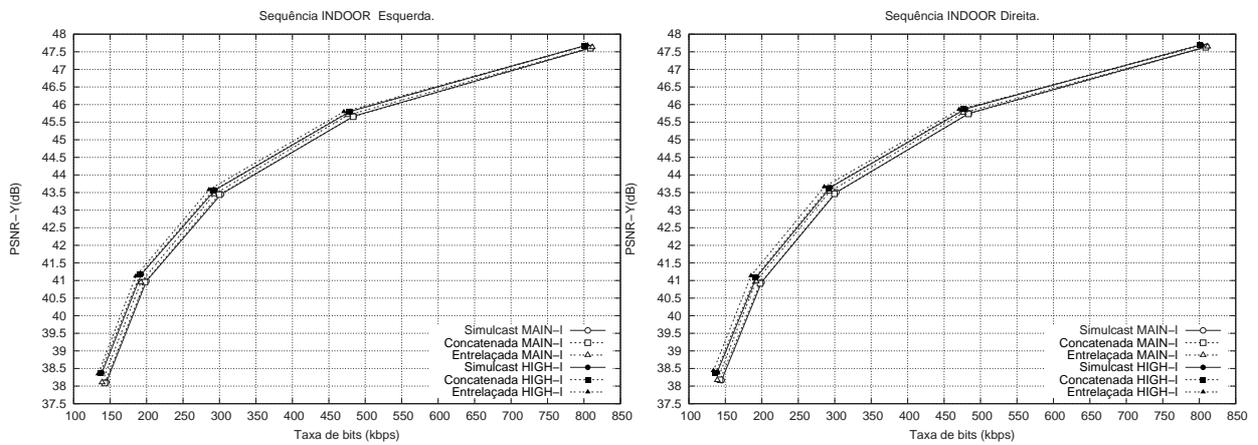
#### 4.5.3 RESULTADOS - SEQUÊNCIA INDOOR.

A FIG. 4.15 apresenta os resultados das simulações dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” para os GoP II e “Entrelaçado” nos perfis *Main* e *High*. O eixo horizontal do gráfico (Taxa de Bits) representa a taxa de codificação total da Sequência binocular.

Pode-se observar que os resultados obtidos para os métodos nos dois perfis apresentaram resultados similares. O pequeno ganho presente nas curvas simuladas no perfil *High* é devido ao fato de que neste perfil há a transformada adaptativa 4x4/8x8 foi habilitada. Com isso, os valores de qualidade das sequências reconstruídas (PSNR-Y) para todas as simulações, foram similares. Os valores da taxa de bits utilizada nos métodos *Simulcast*



(a) Sequência INDOOR esquerda - GoP II. (b) Sequência INDOOR direita - GoP II.



(c) Sequência INDOOR esquerda - GoP III. (d) Sequência INDOOR direita - GoP III.

FIG. 4.15: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil *MainxHigh* - INDOOR.

(taxa de bits da vista esquerda somada à taxa de bits da vista direita) e “Concatenado” (taxa de bits da seqüência “concatenada”) no perfil *Main* e *High* com estrutura de GoP II e III foram similares. As simulações utilizando o GoP III apresentaram resultados superiores na qualidade das seqüências reconstruídas em relação as realizadas com o GoP II, devido o GoP III apresentar um estrutura que apresenta um número menor de quadros tipo B (PQ constante = 30) e de tamanho menor (número maior de quadros tipo I).

O método “Entrelaçado” na média utiliza uma taxa de bits inferior aos demais e seu desempenho é superior aos métodos *Simulcast* e “Concatenado” (TAB. 4.9 - FIG. 4.15).

As TAB. 4.9 e TAB. 4.10 apresentam alguns resultados selecionados obtidos para os três métodos, com valores similares de qualidade das seqüências reconstruídas (PSNR-

Y), dos GoP II e III, para a comparação da taxa de compressão e do fator de excesso de banda. Nas TAB. 4.13 e TAB. 4.14, os métodos (MTD) *Simulcast*, “Concatenado” e “Entrelaçado” são representados pelos algarismo romanos I, II e III.

TAB. 4.13: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP II. - INDOOR

MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	16	47.00	47.03	47.00	47.02	47.00	47.03	115.38	114.50	115.83	115.83	1.007	1.008
<i>High</i>	16	47.06	47.09	47.06	47.07	47.06	47.10	117.19	115.38	116.73	117.19	1.004	1.008
<i>Main</i>	21	45.31	45.38	45.30	45.37	45.33	45.40	185.19	182.93	184.05	188.68	1.000	1.025
<i>High</i>	21	45.44	45.51	45.42	45.49	45.42	45.50	185.19	185.19	186.34	189.87	1.006	1.025
<i>Main</i>	26	43.28	43.34	43.26	43.32	43.29	43.36	277.78	272.73	280.37	288.46	1.019	1.048
<i>High</i>	26	43.40	43.46	43.39	43.46	43.41	43.47	283.02	283.02	288.46	297.03	1.019	1.050
<i>Main</i>	31	40.91	40.91	40.89	40.88	40.86	40.93	394.74	384.62	400.00	410.96	1.027	1.055
<i>High</i>	31	41.12	41.04	41.11	41.03	41.07	41.05	405.41	405.41	410.96	428.57	1.014	1.057
<i>Main</i>	36	38.15	38.22	38.13	38.21	38.13	38.23	483.87	483.87	491.80	517.24	1.016	1.069
<i>High</i>	36	38.42	38.46	38.41	38.44	38.47	38.49	500.00	500.00	517.24	526.32	1.035	1.053

Podemos observar nas TAB. 4.13 e TAB. 4.14 para valores de PSNR-Y similares, que o método “Entrelaçado” (III) apresenta na média a maior taxa de codificação (TC), tendo o método “Concatenado” (II) o segundo melhor desempenho. E analisando os resultados obtidos em relação ao fator de uso de banda, verificamos que os métodos “Entrelaçado” e “Concatenado” geram trens de bits menores em relação ao método *Simulcast* (I).

#### 4.5.4 RESULTADOS - SEQUÊNCIA OUTDOOR.

A FIG. 4.16 apresenta os resultados das simulações dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” para os GoP II e “Entrelaçado” nos perfis *Main* e *High*. O eixo horizontal do gráfico (Taxa de Bits) representa a taxa de codificação total da seqüência binocular.

Pode-se observar que os resultados obtidos para os métodos nos dois perfis apresentaram resultados similares. O pequeno ganho presente nas curvas simuladas no perfil *High* é devido ao fato de que neste perfil há a transformada adaptativa 4x4/8x8 foi habilitada. Com isso, os valores de qualidade das seqüências reconstruídas (PSNR-Y) para todas as

TAB. 4.14: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP III. - INDOOR

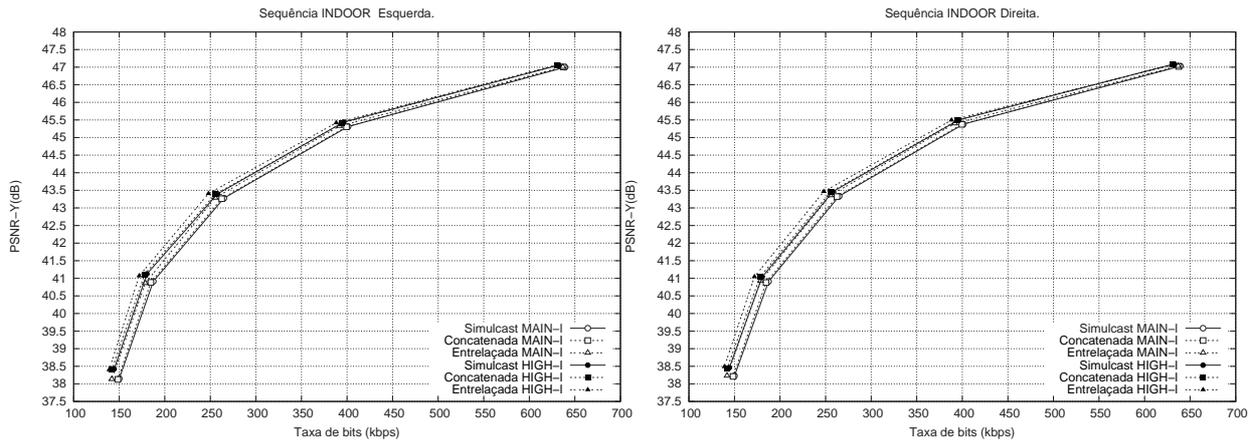
MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	16	47.61	47.65	47.60	47.63	47.63	47.65	92.02	91.46	92.02	91.74	1.003	1.000
<i>High</i>	16	47.68	47.69	47.66	47.69	47.69	47.69	93.17	92.59	93.17	92.59	1.003	0.997
<i>Main</i>	21	45.67	45.75	45.65	45.74	45.70	45.76	154.64	153.06	153.85	156.25	1.000	1.042
<i>High</i>	21	45.81	45.88	45.79	45.87	45.80	45.86	154.64	154.64	156.25	157.89	1.010	1.021
<i>Main</i>	26	43.45	43.50	43.44	43.46	43.45	43.52	245.90	241.94	247.93	254.24	1.017	1.039
<i>High</i>	26	43.57	43.65	43.55	43.62	43.58	43.66	254.24	250.00	254.24	260.87	1.009	1.035
<i>Main</i>	31	40.98	40.96	40.95	40.92	40.95	40.98	375.00	365.85	375.00	384.62	1.013	1.039
<i>High</i>	31	41.18	41.10	41.16	41.09	41.13	41.14	384.62	384.62	389.61	400.00	1.014	1.040
<i>Main</i>	36	38.11	38.17	38.09	38.18	38.11	38.17	517.24	500.00	517.24	535.71	1.0186	1.054
<i>High</i>	36	38.38	38.39	38.36	38.35	38.35	38.44	537.71	517.24	545.45	555.56	1.036	1.056

simulações, foram similares. Os valores da taxa de bits utilizada nos métodos *Simulcast* (taxa de bits da vista esquerda somada à taxa de bits da vista direita) e “Concatenado” (taxa de bits da seqüência “concatenada”) no perfil *Main* e *High* com estrutura de GoP II e III foram similares. As simulações utilizando o GoP III apresentaram resultados superiores na qualidade das seqüências reconstruídas em relação as realizadas com o GoP II, devido o GoP III apresentar um estrutura que apresenta um número menor de quadros tipo B (PQ constante = 30) e de tamanho menor (número maior de quadros tipo I).

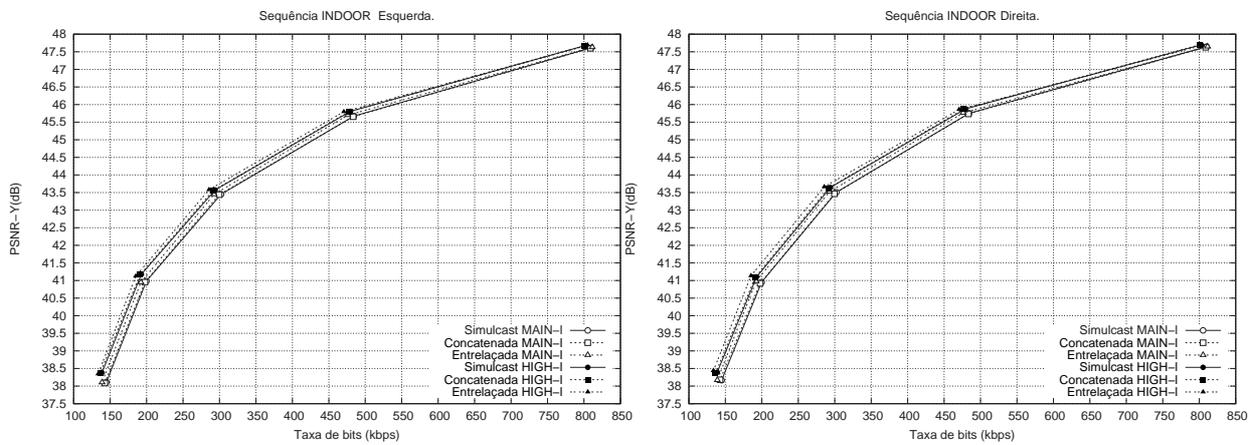
O método “Entrelaçado” na média utiliza uma taxa de bits inferior aos demais e seu desempenho é superior aos métodos *Simulcast* e “Concatenado” (TAB. 4.9 - FIG. 4.16).

As TAB. 4.15 e TAB. 4.16 apresentam alguns resultados selecionados obtidos para os três métodos, com valores similares de qualidade das seqüências reconstruídas (PSNR-Y), dos GoP II e III, para a comparação da taxa de compressão e do fator de excesso de banda. Nas TAB. 4.15 e TAB. 4.16, os métodos (MTD) *Simulcast*, “Concatenado” e “Entrelaçado” são representados pelos algarismo romanos I, II e III.

Podemos observar nas TAB. 4.15 e TAB. 4.16 para valores de PSNR-Y similares, que o método “Entrelaçado” (III) apresenta na média a maior taxa de codificação (TC), tendo o método “Concatenado” (II) o segundo melhor desempenho. E analisando os resultados obtidos em relação ao fator de uso de banda, verificamos que os métodos “Entrelaçado” e “Concatenado” geram trens de bits menores em relação ao método *Simulcast* (I).



(a) Seqüência OUTDOOR esquerda - GoP II. (b) Seqüência OUTDOOR direita - GoP II.



(c) Seqüência OUTDOOR esquerda - GoP III (d) Seqüência OUTDOOR direita - GoP III.

FIG. 4.16: Comparação dos métodos *Simulcast*, “Concatenado” e “Entrelaçado”. - GoP II e III - Perfil *MainxHigh* - OUTDOOR

#### 4.6 COMPARAÇÃO ENTRE OS MÉTODOS - MÉTODO “BINOCULAR”

As seqüências foram simulados no perfil *Main*. O método “Entrelaçado” foi codificado com a *flag IntraBotton* habilitada, a qual habilita a codificação intra-quadro, tanto para o campo par quanto para o ímpar. Os parâmetros de quantização dos mapas de disparidades e dos resíduos do compensador de disparidades serão apresentados para cada seqüência simulada. O método binocular tem seu desempenho diretamente relacionado com a qualidade do mapa de disparidades gerado no processo de estimação de disparidades. Ele também apresenta uma complexidade maior em relação aos outros métodos,

TAB. 4.15: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP II. - OUTDOOR

MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	16	45.60	45.62	45.60	45.58	45.60	45.61	70.42	69.77	70.09	71.26	1.000	1.017
<i>High</i>	16	45.65	45.65	45.64	45.61	45.64	45.63	71.77	71.09	71.60	72.64	1.002	1.017
<i>Main</i>	21	42.65	43.01	42.93	42.99	42.95	43.00	108.70	107.14	108.30	111.94	1.004	1.037
<i>High</i>	21	43.06	43.12	43.05	43.10	43.06	43.11	111.11	109.49	111.11	114.07	1.007	1.034
<i>Main</i>	26	40.36	40.48	40.34	40.46	40.40	40.49	164.84	161.29	163.93	171.43	1.001	1.051
<i>High</i>	26	40.46	40.55	40.45	40.55	40.52	40.58	172.41	164.54	170.45	177.51	1.000	1.041
<i>Main</i>	31	37.91	37.96	37.89	37.95	37.96	37.98	234.38	230.77	234.38	243.90	1.008	1.049
<i>High</i>	31	38.13	38.19	38.11	38.17	38.20	38.22	245.90	238.10	243.90	254.24	1.008	1.051
<i>Main</i>	36	35.47	35.44	35.44	35.42	35.54	35.50	300.00	288.46	300.00	309.28	1.020	1.052
<i>High</i>	36	35.66	35.74	35.62	35.67	35.76	35.80	312.50	249.12	312.50	315.79	1.031	1.042

TAB. 4.16: Comparação - Métodos *Simulcast*, “Concatenado” e “Entrelaçado”. Perfil *Main* e *High*. GoP III. - OUTDOOR

MTD		I		II		III		TC				FEB	
Perfil	PQ	PSNR		PSNR		PSNR		I	I	II	III	I/II	I/III
-	-	E(dB)	D(dB)	E(dB)	D(dB)	E(dB)	D(dB)	E	D	ED	ED	-	-
<i>Main</i>	16	46.01	46.03	45.99	46.01	46.01	46.01	56.82	56.18	56.60	57.25	1.002	1.013
<i>High</i>	16	46.06	46.08	46.05	46.30	46.03	46.04	57.92	57.47	57.80	58.48	1.002	1.014
<i>Main</i>	21	43.14	43.21	43.14	43.19	43.13	43.20	89.82	88.76	89.82	92.59	1.006	1.037
<i>High</i>	21	43.25	43.33	43.23	43.30	43.24	43.31	92.02	90.91	91.74	94.04	1.003	1.028
<i>Main</i>	26	40.41	40.53	40.39	40.51	40.46	40.56	141.51	138.89	140.85	147.06	1.005	1.050
<i>High</i>	26	40.53	40.63	40.52	40.61	40.58	40.65	147.06	144.23	147.06	153.06	1.010	1.051
<i>Main</i>	31	37.88	37.93	37.86	37.90	37.92	37.93	214.29	208.33	212.77	222.22	1.007	1.052
<i>High</i>	31	38.09	38.14	38.07	38.13	38.16	38.19	223.88	217.39	223.88	232.56	1.015	1.054
<i>Main</i>	36	35.34	35.32	35.31	35.29	35.42	35.39	300.00	288.46	300.00	309.28	1.020	1.052
<i>High</i>	36	35.49	35.58	35.49	35.56	35.56	35.63	312.50	300.00	309.28	322.58	1.010	1.054

pois além ter um processamento extra na estimação de disparidades dos pares estéreo, o mesmo gera três trens de bits, precisando de uma sincronização dos mesmos no decodificador, reconstrução do par estéreo e armazenamento. No processo de reconstrução da vista direita foi utilizado um interpolador linear para minimizar as falhas geradas pelo mapa de disparidades. Com isso, foi obtido um RCD mais homogêneo. É importante e esperado para um melhor desempenho no processo de recuperação de profundidade da

cena 3D, que a vista direita reconstruída tenha qualidade similar da vista esquerda codificada/decodificada, utilizada no processo de estimação/compensação de disparidades. A título de esclarecimento, foi efetuado um teste para mostrar que o método é eficiente quando o mapa de disparidades tem alta qualidade. Nas figuras seguintes, os valores das intensidades dos pixels dos mapas de disparidades correspondem aos valores das disparidades. As disparidades são representadas em uma escala de cinza, sendo a menor disparidade representada pelo nível preto e a disparidade máxima pelo branco. Já nos mapas de oclusões, os pixels em branco representam as oclusões encontradas na estimação de disparidades.

#### 4.6.1 RESULTADOS - SEQUÊNCIA MAN.

Os tipos de GoPs e os parâmetros de quantização utilizados para a seqüência MAN foram: GoP II e PQ {00, 04, 10, 16, 22, 28, 34} (FIG. 4.17a e FIG. 4.17b), para o mapa de disparidades e resíduo do compensador de disparidade.

O RCD foi gerado a partir da seqüência esquerda codificada com o PQ de valor 16 e três mapas de disparidades codificados com os PQs {00, 22, 34}, com o objetivo de avaliar a vista direita reconstruída com diferentes valores de PSNR-Y dos mapas e RCDs.

As FIG. 4.17a e FIG. 4.17b apresenta o mapa de disparidades e de oclusão do par estéreo 25 da seqüência MAN, e os resultados das simulações do método “Binocular” com o GoP II no perfil *Main*. Podemos observar que o estimador de disparidades é ineficiente nas áreas homogêneas da imagem (plano de fundo e alguns pontos no rosto) e também que o mapa é formado por algumas áreas homogêneas.

Podemos observar que a seqüência MAN apresenta grandes áreas de oclusão provocadas pela acentuada distorção perspectiva presente na mesma (FIG. 4.17b) e por erros no processo de estimação (canto superior direito do mapa de oclusões (FIG. 4.17b)). Com isso, no processo de geração do RCD, o mesmo apresenta variações bruscas na intensidade dos seus pixels (alta freqüência), resultando numa codificação ineficiente.

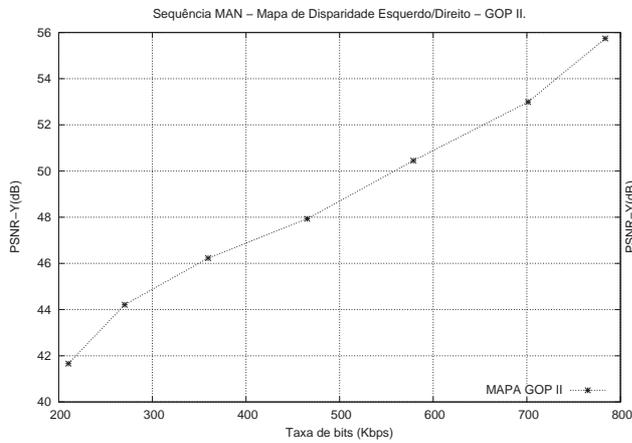
Na FIG. 4.17e, o eixo horizontal do gráfico (Taxa de Bits) representa para o método “Binocular”, à taxa de codificação utilizada na codificação do mapa de disparidades mais a do RCD. Esta figura apresenta os resultados das simulações do método “Binocular” com o GoP II no perfil *Main* da seqüência direita reconstruída a partir de três mapas de disparidades em relação à seqüência direita codificada através do método *Simulcast* com



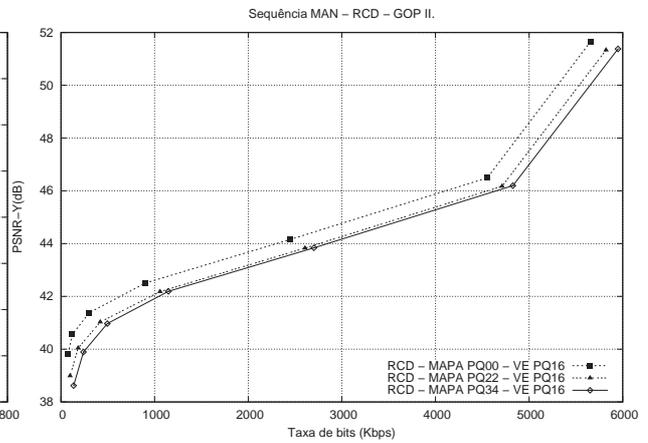
(a) Mapa de Disparidades E/D.



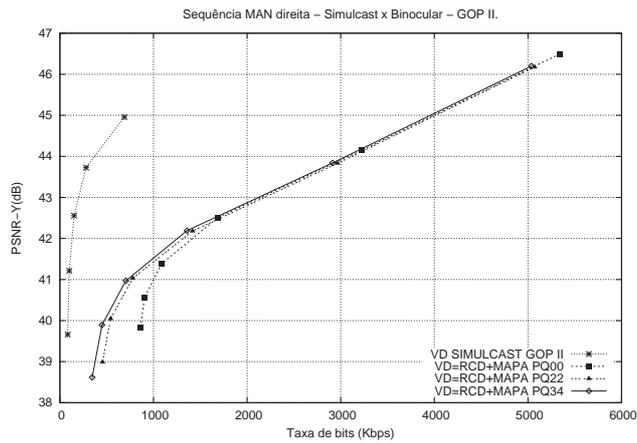
(b) Mapa de Oclusões E/D.



(c) Mapa de Disparidades E/D - GoP II.



(d) RCD - GoP II.



(e) Sequência MAN direita reconstruída - GoP II.

FIG. 4.17: Comparação dos métodos *Simulcast* e "Binocular". - Sequência MAN - Perfil *Main*.

## GoP II.

Quando se codifica uma seqüência estéreo é importante, principalmente para o processo de recuperação da profundidade da cena 3D, que as vistas do par estéreo codificado apresente uma qualidade similar. Como no método *Simulcast* as vistas esquerda e direita codificadas apresentam valores de PSNR-Y similares, a vista direita deste método com o mesmo PQ usado na vista esquerda utilizada no método “Binocular”, serviu como referência para avaliar o desempenho da vista direita reconstruído do método “Binocular”. Na FIG. 4.17e, vemos que o método “Binocular”, para alcançar o mesmo valor de PSNR (45,60dB) da seqüência direita reconstruída no método *Simulcast* (TAB. 4.9), utiliza uma taxa extra de codificação de aproximadamente 3130 Kbps. E mesmo para o pior desempenho do método *Simulcast* com essa configuração, o método “Binocular” utiliza uma taxa extra de codificação de aproximadamente 250 Kbps. Essa diferença observada entre as curvas dos métodos *Simulcast* e “Binocular” é atribuída a codificação do RCD, pois o estimador de disparidades gerou um mapa de disparidades ineficiente que resultou em falhas na reconstrução da vista direita. Com isso, na compensação de disparidades, o RCD gerado apresentou várias regiões (correspondentes a regiões na cabeça e camisa) com variações bruscas na intensidade dos pixels, favorecendo uma codificação ineficiente, aumentando assim o número de bits utilizados na codificação de cada quadro do RCD.

### 4.6.2 RESULTADOS - SEQÜÊNCIA SAXO.

Os tipos de GoPs e os parâmetros de quantização utilizados para a seqüência SAXO foram: GoP II e PQs {00, 04, 10, 16, 21, 26, 28, 31, 34} para o mapa de disparidades e GoP II, PQ {00, 04, 10, 16, 22, 28, 34, 40} para o resíduo do compensador de disparidade (FIG. 4.18a e FIG. 4.18b).

O RCD foi gerado a partir de seqüência esquerda codificada com o PQ de valor 21, combinado com três mapas de disparidades codificados com os PQs {00, 22, 34} com o objetivo de avaliar a vista direita reconstruída com diferentes valores de PSNR-Y dos mapas e RCDs.

A FIG. 4.18 apresentam o mapa de disparidades e de oclusão do par estéreo 25 da seqüência, e os resultados das simulação do método “Binocular” com o GoP II no perfil *Main*. Podemos observar na FIG. 4.18a que o estimador de disparidades é ineficiente na geração do mapa de disparidades. O mapa apresenta variações dos valores de disparidades

em várias regiões da imagem com a mesma profundidade. Essa característica diminui o desempenho da codificação dos mapas e gera no compensador de disparidades uma vista direita reconstruída com falhas. A FIG. 4.18b mostra o mapa de oclusão gerado pelo estimador de disparidades. Podemos observar que o mesmo apresenta grandes áreas de oclusão (nível branco). Com isso, a vista direita reconstruída apresenta falhas e variações na intensidade de alguns pixels (gerada pela interpolação linear) comparados a vista direita original, resultando em um RCD com variações bruscas na intensidade dos seus pixels (alta frequência) principalmente nas regiões dentro do corpo do saxofonista e no antebraço do outro homem da cena. Este tipo de seqüência apresenta uma codificação ineficiente.

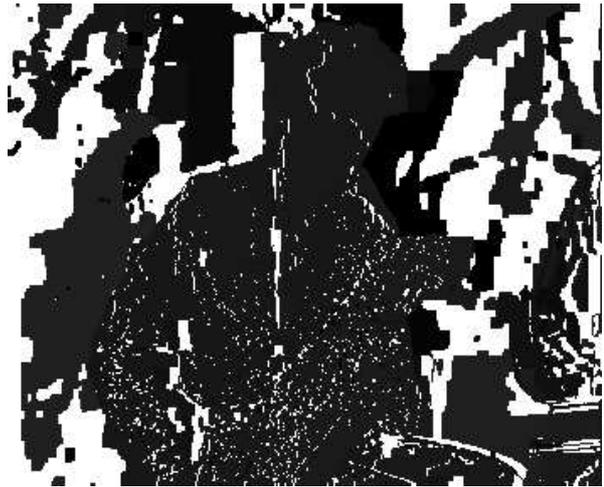
Na FIG. 4.18e, o eixo horizontal do gráfico (Taxa de Bits) representa para o método “Binocular” à taxa de codificação da mapa de disparidades mais a do RCD. A FIG. 4.18e apresenta os resultados das simulações do método “Binocular” com o GoP II no perfil *Main* da seqüência direita reconstruída a partir de três mapas de disparidades, em relação a seqüência direita codificada através do método *Simulcast* com GoP II.

Quando se codifica uma seqüência estéreo é importante, principalmente para o processo de recuperação da profundidade da cena 3D, que as vistas do par estéreo codificado apresente uma qualidade similar. Como no método *Simulcast* as vistas esquerda e direita codificadas apresentam valores de PSNR-Y similares, a vista direita deste método com o mesmo PQ usado na vista esquerda utilizada no método “Binocular”, serviu como referência para avaliar o desempenho da vista direita reconstruído do método “Binocular”. Na FIG. 4.18e, vemos que o método “Binocular” para alcançar o mesmo valor de PSNR (37,37 dB) da seqüência direita reconstruída no método *Simulcast* com parâmetro de quantização de valor 21 e com GoP II (TAB. 4.11), utiliza uma taxa extra de codificação de aproximadamente 1620 Kbps que é aproximadamente à taxa gasta na codificação da seqüência direita. E, para o pior caso do método *Simulcast* (PSNR-Y = 33,87 dB), com essa configuração, o método “Binocular” utiliza uma taxa extra de codificação de aproximadamente 800 Kbps. A diferença entre as curvas dos métodos *Simulcast* e “Binocular” é gerada principalmente pela codificação ineficiente do RCD, a qual resulta em uma utilização excessiva de bits por quadro.

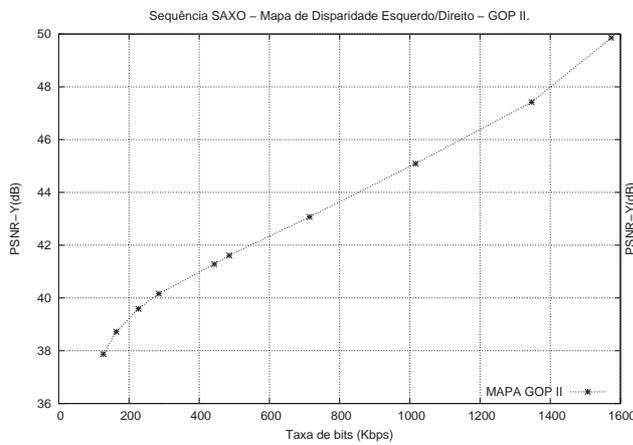
#### 4.6.3 RESULTADOS - SEQÜÊNCIA INDOOR.



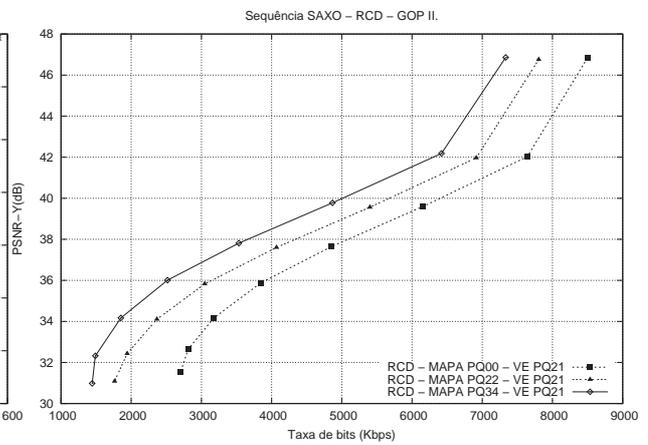
(a) Mapa de Disparidades



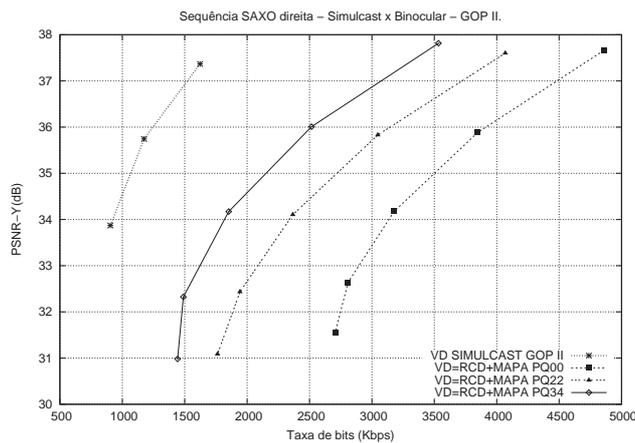
(b) Mapa de Oclusões



(c) Mapa de Disparidades E/D - GoP II



(d) RCD - GoP II



(e) Sequência SAXO direita reconstruída - GoP II

FIG. 4.18: Comparação dos métodos *Simulcast* e “Binocular”. - Sequência SAXO - Perfil *Main*

Os tipos de GoPs e os parâmetros de quantização utilizados para a seqüência INDOOR foram: mapa de disparidades com GoP II e PQs {00, 04, 10, 16, 22, 28, 30, 32, 34} e Resíduo do Compensador de Disparidade com GoP II, PQ {00, 04, 10, 16, 22, 28, 34} (FIG. 4.19a e FIG. 4.19b).

O RCD foi gerado a partir de seqüência esquerda codificada com o PQ de valor 16 com GoP II e três mapas de disparidades codificados com os PQs {00, 22, 34}, com o objetivo de avaliar a vista direita reconstruída com diferentes valores de PSNR-Y dos mapas e RCDs.

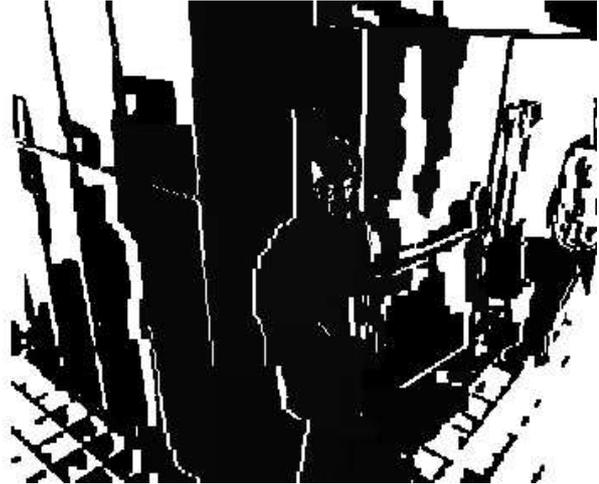
Na FIG. 4.19a, podemos observar que o estimador de disparidades apresenta falhas no processo de estimação, atribuindo valores de disparidades elevados à regiões de baixa disparidades (pixels em branco). O mapa de disparidades apresenta grandes áreas homogêneas, facilitando o processo de codificação. Já a FIG. 4.19b confirmar a deficiência do estimador de disparidades em áreas homogêneas. O mesmo marcar oclusões em áreas visíveis nas duas imagens.

Na FIG. 4.19e, o eixo horizontal do gráfico (Taxa de Bits) representa para o método “Binocular”, à taxa de codificação do mapa de disparidades e do RCD. A FIG. 4.19e apresenta os resultados das simulações do método “Binocular” com o GoP II no perfil *Main* da seqüência direita reconstruída a partir de três mapas de disparidades em relação à seqüência direita codificada através do método *Simulcast* com GoP II.

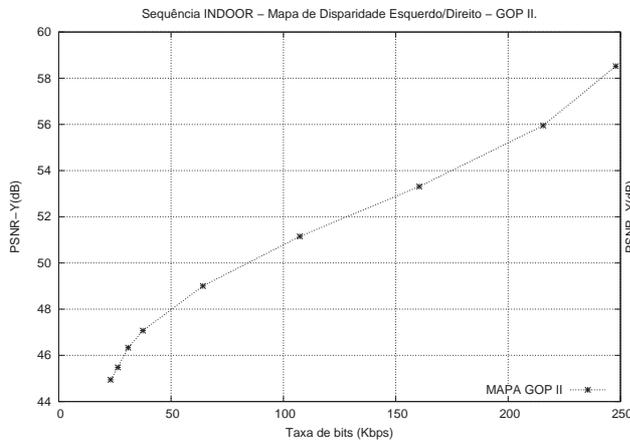
Quando se codifica uma seqüência estéreo é importante, principalmente para o processo de recuperação da profundidade da cena 3D, que as vistas do par estéreo codificado apresente uma qualidade similar. Como no método *Simulcast* as vistas esquerda e direita codificadas apresentam valores de PSNR-Y similares, a vista direita deste método com o mesmo PQ usado na vista esquerda utilizada no método “Binocular”, serviu como referência para avaliar o desempenho da vista direita reconstruído do método “Binocular”. Na FIG. 4.19c, vemos que o método “Binocular” para alcançar o mesmo valor de PSNR (47,03 dB) da seqüência direita reconstruída no método *Simulcast* com parâmetro de quantização de valor 16 e com GoP II (TAB. 4.13), utiliza uma taxa extra de codificação de aproximadamente 1620 Kbps que é aproximadamente a mesma taxa gasta na codificação da seqüência direita. E para um valor de PSNR-Y igual 40,0 dB o método *Simulcast* e “Binocular” utilizam a mesma taxa de codificação. As características da seqüência (aproximadamente quarenta e um quadros estáticos e sessenta com movimento), do mapa de disparidades (grandes áreas homogêneas) e do RCD (grandes áreas homogêneas nos



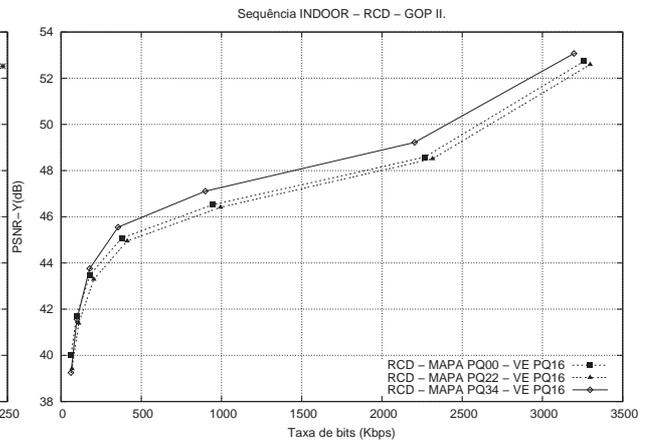
(a) Mapa de Disparidades



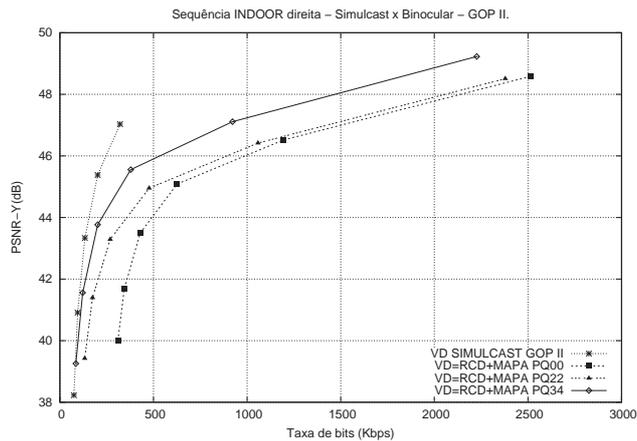
(b) Mapa de Oclusões



(c) Mapa de Disparidades E/D - GoP II



(d) RCD - GoP II



(e) Sequência INDOOR direita reconstruída - GoP II

FIG. 4.19: Comparação dos métodos *Simulcast* e “Binocular”. - Sequência INDOOR - Perfil *Main*

quadros estáticos) possibilitaram o desempenho similar do método “Binocular” para o valor de PSNR-Y igual 40,0 dB em relação ao método *Simulcast*.

#### 4.6.4 RESULTADOS - SEQUÊNCIA OUTDOOR.

Os tipos de GoPs e os parâmetros de quantização utilizados para a sequência OUTDOOR foram: mapa de disparidades com GoP II e PQs {00, 04, 10, 16, 21, 26, 28, 31, 34} e Resíduo do Compensador de Disparidade com GoP II, PQ {00, 04, 10, 16, 22, 28, 34, 40} (FIG. 4.20a e FIG. 4.20b).

O RCD foi gerado a partir de sequência esquerda codificada com o PQ de valor 21 com GoP II e três mapas de disparidades codificados com os PQs {00, 22, 34}, com o objetivo de avaliar a vista direita reconstruída com diferentes valores de PSNR-Y dos mapas e RCDs.

Na FIG. 4.20a, podemos observar que o estimador de disparidades apresenta falhas na geração do mapa de disparidades, atribuindo valores de disparidades elevados para regiões com valores de disparidades pequenos. Já a FIG. 4.20b mostra que o estimador de disparidades falha na determinação das oclusões.

Na FIG. 4.20e, o eixo horizontal do gráfico (Taxa de Bits) representa para o método binocular, à taxa de codificação do mapa de disparidades e do RCD. A FIG. 4.20e apresenta os resultados das simulações do método “Binocular” com o GoP II no perfil *Main* da sequência direita reconstruída a partir de três mapas de disparidades, em relação à sequência direita codificada através do método *Simulcast* com GoP II.

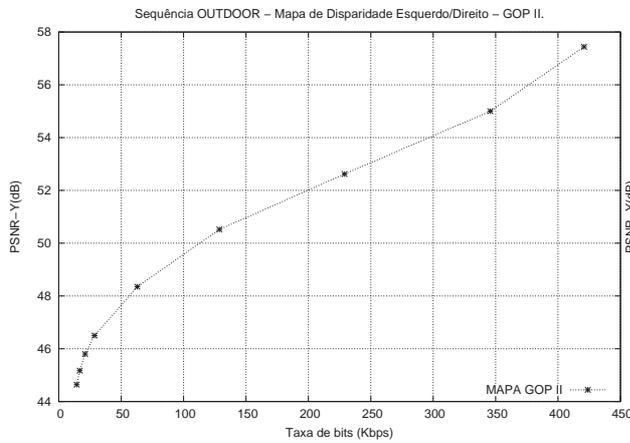
Quando se codifica uma sequência estéreo é importante, principalmente para o processo de recuperação da profundidade da cena 3D, que as vistas do par estéreo codificado apresente uma qualidade similar. Como no método *Simulcast* as vistas esquerda e direita codificadas apresentam valores de PSNR-Y similares, a vista direita deste método com o mesmo PQ usado na vista esquerda utilizada no método “Binocular”, serviu como referência para avaliar o desempenho da vista direita reconstruído do método “Binocular”. Na FIG. 4.20e, vemos que o método “Binocular”, para alcançar o mesmo valor de PSNR (45,62 dB) da sequência direita reconstruída no método *Simulcast* com parâmetro de quantização de valor 16 e com GoP II (TAB. 4.11), utiliza uma taxa extra de codificação de aproximadamente 160 Kbps. E, para um valor de PSNR-Y igual a 38,13 dB, o método “Binocular” utiliza uma taxa de codificação inferior de aproximadamente 46



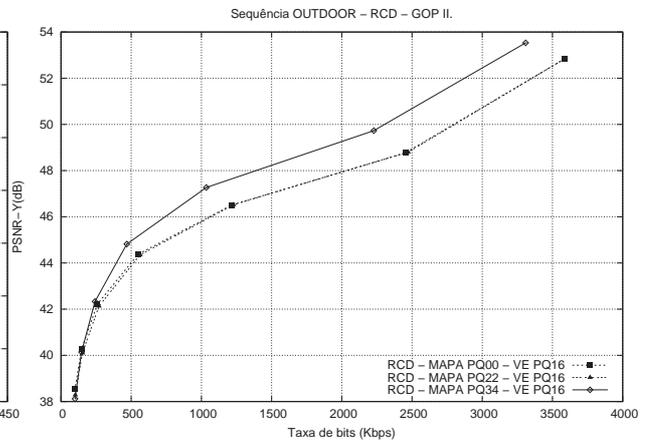
(a) Mapa de Disparidades



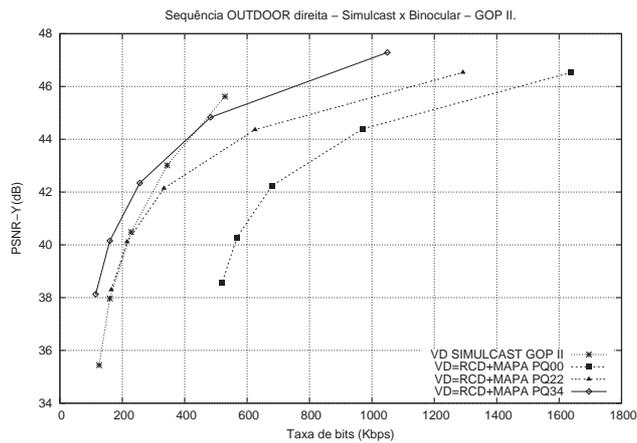
(b) Mapa de Oclusões



(c) Mapa de Disparidades E/D - GoP II



(d) RCD - GoP II



(e) Sequência OUTDOOR direita reconstruída - GoP II

FIG. 4.20: Comparação dos métodos *Simulcast* e “Binocular”. - Sequência OUTDOOR - Perfil *Main*

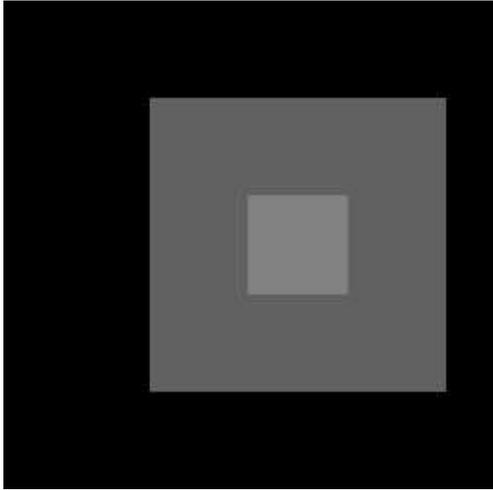
Kbps, comparado ao método *Simulcast*. As características da seqüência (aproximadamente cinquenta e quatro quadros estáticos e quarenta e seis com movimento), do mapa de disparidades (grandes áreas homogêneas) e do RCD (grandes áreas homogêneas nos quadros estáticos) possibilitaram o desempenho superior do método “Binocular” para o valor de PSNR-Y igual 38,13 dB em relação ao método *Simulcast*.

#### 4.6.5 RESULTADOS - SEQÜÊNCIA RANDOT.

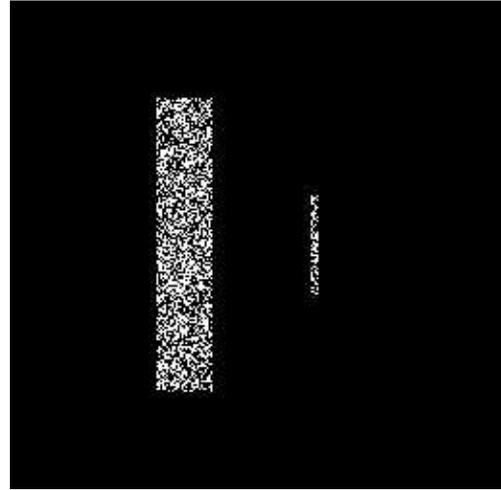
Essa seqüência foi artificialmente gerada. Cada par estéreo da seqüência possui um mapa de disparidades com *Ground Truth* (disparidades conhecidas). Isto é na realidade um caso impossível de se obter, pois necessitaria de um estimador de disparidades perfeito. No entanto, esse seqüência foi gerada para mostrar o desempenho do método “Binocular” fazendo uso de mapas de disparidades “perfeitos”. Conforme será mostrado, o desempenho do método quando há uma geração de um mapa de disparidades com uma acurácia alta é superior aos demais métodos avaliados nesta tese.

A FIG. 4.21 apresenta o mapa de disparidades (Ground Truth) e o RCD, e os resultados das simulações dos métodos *Simulcast*, “Concatenado” e “Entrelaçado” da seqüência RANDOT, com o GoP II no perfil *Main*. O eixo horizontal do gráfico (Taxa de Bits) representa para o método “Binocular”, à taxa de codificação do mapa de disparidades e do RCD.

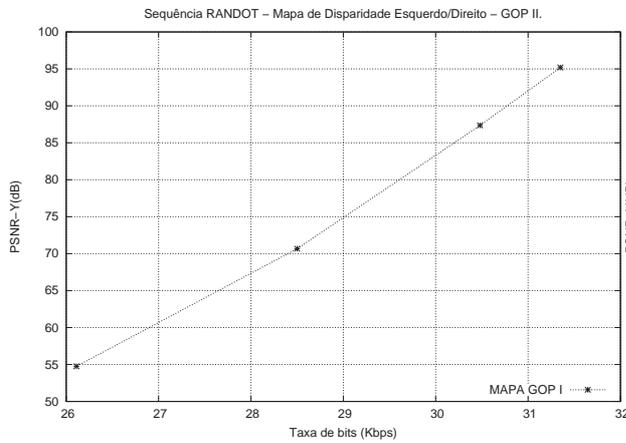
Podemos observar na FIG. 4.21e que o método binocular apresenta o melhor desempenho. Vemos que para um valor de PSNR-Y igual a 50,0 dB, o método “Binocular” (MAPA PQ 00) apresenta uma economia na taxa de codificação em relação ao método “Entrelaçado” de aproximadamente 1500 Kbps e para uma taxa fixa de 2500 Kbps, o mesmo apresenta ganhos de aproximadamente 7,0 dB (MAPA PQ 04) e 12,0 dB (MAPA PQ 00). Isto se deve ao fato de que a seqüência RANDOT possui *Ground Truth* com grandes regiões homogêneas. Uma vez que o mapa de disparidades gerado corresponde com exatidão às disparidades presentes nas duas vistas, a reconstrução da vista direita através da combinação do mapa de disparidades e da vista esquerda apresenta imperfeições somente nas áreas correspondentes a oclusões. Com isso, o RCD FIG. 4.21b será constituído puramente pelas oclusão entre as vistas. Como o RCD apresenta muitas áreas homogêneas, o codificador irá obter uma excelente taxa de compressão. Outro fator que aumenta o desempenho do método “Binocular” é que o mapa de disparidades da



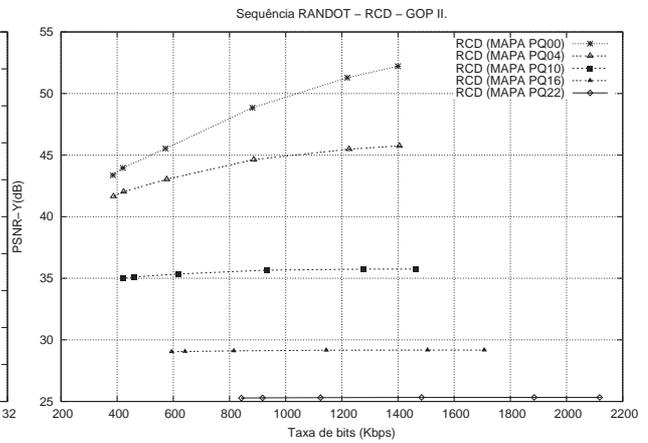
(a) Mapa de Disparidades



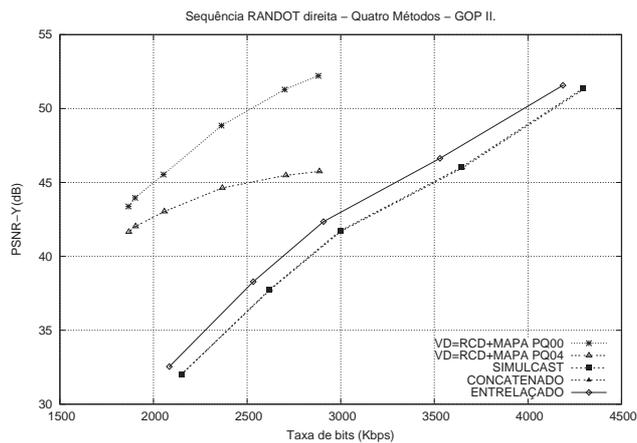
(b) RCD



(c) Mapa de Disparidades E/D - GoP II



(d) RCD - GoP II



(e) Sequência RANDOT direita reconstruída - GoP II

FIG. 4.21: Comparação dos métodos *Simulcast*, “Concatenado”, “Entrelaçado” e “Bino-  
cular”. - Sequência RANDOT - Perfil *Main*

seqüência RANDOT apresenta grandes áreas homogêneas. Neste caso, o método “Entrelaçado” apresentou o segundo melhor desempenho. (FIG. 4.21e)

## 4.7 RESUMO

Neste capítulo, foi avaliado o desempenho de quatro métodos de codificação de vídeo estéreo.

Após várias simulações com um grande número de parâmetros de configuração, constatou-se que o método “Entrelaçado” apresenta o melhor desempenho entre os métodos avaliados sendo o mesmo indicado para codificação de vídeo estéreo. O ganho na codificação utilizando o método “Entrelaçado” é obtido principalmente da predição das vistas esquerda e direita do quadro entrelaçado para pares estéreos com valores de disparidades baixos. Porém foi observado que para seqüências estéreo com valores de disparidades altos, o método não decidiu, na maioria dos testes, pela predição entre as vistas esquerda e direita do próprio quadro entrelaçado quando foi utilizado um único quadro de referência. Os testes indicam que, quando a seqüência estéreo possui uma disparidade grande (FIG. 4.2a), a predição é realizada com vistas esquerdas e direitas dos quadros anteriores e posteriores (quadro tipo B). Por isso, quando utilizamos um número de quadros de referência grande, o desempenho da seqüência com um valor disparidade elevado alcançava resultados similares para os dois modos de codificação avaliados em (SUN, 2005) e para os três (método “Entrelaçado”, “Entrelaçado” com a ferramenta PAFF e “Entrelaçado” com as ferramenta PAFF + MBAFF) testados nesta dissertação. Uma conclusão mais precisa não foi possível devido à disponibilidade de uma única seqüência estéreo com esta característica. Os resultados obtidos nas simulações com taxas de bits constante indicam que o método “Entrelaçado” tenderá a manter seu desempenho superior aos demais métodos para outras taxas de bits. O método “Concatenado” obteve o segundo melhor desempenho. Já o método *Simulcast* apresentou em muitas simulações taxas de codificação superiores em relação aos outros métodos, em alguns casos se igualando com o método “Concatenado”. Já nas simulações com taxas de bits variáveis, o desempenho do método “Entrelaçado foi similar aos métodos “Concatenado” e *Simulcast* pelo fato de que na codificação dos quadro tipo B da seqüência, o número de quadros de referência posteriores foi fixado em um. Desta forma, afetando o desempenho do método “Entrelaçado. O método “Binocular” apresentou o pior desempenho entre os métodos avaliados, pois o mesmo depende diretamente da qualidade do mapa de disparidades. A título ilustra-

tivo, foi mostrado que este método, quando utiliza um mapa de disparidades conhecidas *Ground Truth*, apresenta um desempenho superior aos demais métodos avaliados. Porém, não é possível ter uma conclusão mais precisa, pois a seqüência RANDOT utilizada no teste foi gerada da imagem artificial estéreo RANDOT, sendo impossível uma seqüência natural possuir as mesma características. Ainda assim, é necessário mais testes com outras seqüências com *Ground Truth* para uma avaliação mais precisa do desempenho do método.

## 5 CONCLUSÕES E COMENTÁRIOS FINAIS

### 5.1 CONCLUSÕES

Nesta dissertação, foram estudados, avaliados e comparados quatro métodos de codificação de vídeo estéreo, sendo que dois destes foram propostos em (SUN, 2003) como suporte à codificação de vídeo estéreo do codificador H.264/AVC.

Os resultados mostraram que, na codificação com taxas de bits constante, o método “Entrelaçado” apresenta resultados superiores aos métodos “Concatenado” e *Simulcast*, e que este necessita de uma taxa de codificação maior para um mesmo valor de qualidade da seqüência reconstruída (PSNR-Y). O método *Simulcast* também aumenta a complexidade no manuseio do trem de bits, por necessitar de sincronização dos quadros no decodificador, ao contrário dos métodos “Concatenado” e “Entrelaçado” que utilizam na codificação da seqüência binocular codificadores tipicamente monoculares. Os resultados das simulações indicam que o método “Entrelaçado” manterá o desempenho alcançado em relação aos outros dois métodos para outras taxas de codificação constantes.

Na codificação com taxa de bits variáveis, os dois perfis (*Main* e *High*) utilizados produziram resultados similares. Os métodos *Simulcast*, “Concatenado” e “Entrelaçado” apresentaram resultados similares. O desempenho inferior do método “Entrelaçado” com taxa de bits variável em relação aos resultados alcançados com taxas de bits constante, é devido ao fato de que na configuração dos parâmetros do codificador, foi estabelecido que na estimação de movimento dos quadros do tipo B seria possível utilizar cinco quadros anteriores e somente um quadro posterior como referência. Desta forma, diminuindo o desempenho do método “Entrelaçado”.

A codificação com taxa de bits variável, em todos os métodos e perfis, produziu valores de PSNR-Y similares em todas as simulações. Os métodos “Concatenado” e “Entrelaçado” apresentaram na maioria das simulações resultados superiores na taxa de compressão comparados ao método *Simulcast*, para uma qualidade de reconstruções das seqüências similar.

Os métodos “Concatenado” e “Entrelaçado” adaptam uma seqüência binocular em

uma única seqüência monocular, diminuindo a complexidade do manuseio do trem de bits.

O ganho na codificação utilizando o método “Entrelaçado” é obtido principalmente da predição da vista direita pela vista esquerda codificada do quadro entrelaçado para seqüências estéreo com valores de disparidade baixos. Foi observado através dos testes que para seqüências estéreo com valores de disparidade de médio a alto, o método não decidiu, na maioria dos testes, em utilizar na predição da vista direita a vista esquerda codificada do próprio quadro entrelaçado quando foi usado um único quadro de referência na estimação de movimento. E que quando se utilizou um valor de quadros de referência alto, a predição das vistas direitas foi realizada com as vistas direitas codificadas, e as vistas esquerdas com as vistas esquerdas codificadas. Desta forma, os testes indicam que quando se pode utilizar muitos quadros como referência na estimação de movimento, o desempenho do método “Entrelaçado” é similar independente se o mesmo está utilizando uma ferramenta de codificação de vídeo entrelaçado (PAFF ou MBAFF) ou não e também para seqüências estéreo com valores de disparidades baixos, médios e altos.

Uma conclusão mais precisa não foi possível devido a disponibilidade de uma única seqüência estéreo com esta característica.

No método “Binocular” há um aumento na complexidade no manuseio do trem de bits por necessitar de sincronização do conjunto quadro, mapa e RCD, no decodificador. O mesmo gera excesso de banda e apresenta uma taxa de codificação na maioria das seqüências simuladas superior aos demais métodos para um mesmo valor de PSNR-Y. O desempenho do método “Binocular” está diretamente relacionado com a acurácia do mapa de disparidades gerado pelo estimador de disparidades. O método “Binocular” apresentou resultados superiores aos demais métodos, quando testado com uma seqüência com o mapa de disparidades “conhecido” (*Ground Truth*). Há a necessidade de mais testes para concluir a superioridade do método “Binocular” aos demais métodos quando a estimação de disparidades do par estéreo possui uma acurácia alta.

## 5.2 TRABALHOS FUTUROS

Estimadores de disparidades com maior acurácia na geração dos mapas de disparidades e codificadores de vídeo estéreo de ótimo desempenho em relação a taxa de codificação e a qualidade das seqüências reconstruídas continuam sendo um desafio a ser investigado. Como também realizar análise mais completa sobre as ferramentas de codificação de

vídeo entrelaçado presentes no codificador H.264/AVC e pesquisar a disponibilidade de seqüências estéreo com grande valores disparidades para a realização de mais testes. Com base nisto, propomos os seguintes tópicos para pesquisas futuras:

- Investigar a possibilidade de inserir os métodos dois, três e quatro dentro do código do codificador H.264/AVC, em conjunto com a mensagem *Supplemental Enhancement Information* (SEI);
- Investigar o desempenho dos métodos avaliados para seqüências estéreo com um valor de disparidade elevada;
- Investigar o uso de outras seqüências estéreo e a disponibilidade destas, com características diversas de textura, movimento, disparidade e também seqüências com *Ground Truth*;

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- ACCAME, M., NATALE, F. G. B. D. e GIUSTO, D. **Hierarchical block matching for disparity estimation in stereo sequences.** *In: Proceedings of the 1995 IEEE International Conference on Image Processing, ICIP'95*, págs. 374–377, 1995.
- ANTONINI, M., B. M. M. P. e DAUBECHIES, I. **Image coding using wavelet transform.** *In: IEEE Trans. Image Proc.*, 1:205–221, 1992.
- BARNARD, S. T. e FISCHLER, M. A. **Computational stereo.** *In: ACM Computing Surveys*, 14:553–572, 1982.
- BHAT, D. N. e NAYAR, S. K. **Ordinal measures for image correspondence.** *In: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:415–423, 1998.
- BIRCHFIELD, S. e TOMASI, C. **Multiway cut for stereo and motion with slanted surfaces.** *In: Proceedings of the 1999 IEEE International Conference on Computer Vision, ICCV'99*, págs. 489–495, 1999.
- BOYKOV, Y., VEKSLER, O. e ZABIH., R. **Markov random fields with efficient approximations.** *In: IEEE Conference on Computer Vision and Pattern Recognition*, págs. 648–655, 1998.
- BOYKOV, Y., VEKSLER, O. e ZABIH., R. **Fast approximate energy minimization via graph cuts.** *In: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, novembro 2001.
- BROW, M. Z., BURSCHKA, D. e HAGER, G. D. **Advances in computational stereo.** *In: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:993–1008, 2003.
- CCETT. **Sequências de teste captadas e distribuídas pelos projetos europeus RACE-DISTIMA e ACTS-PANORAMA, França.** *In: Centre Commun d'Etudes de Télédiffusion et Télécommunications.*
- CORMEN, T. H., LEISERSON, C. E. e RIVEST, R. L. **Introduction to Algorithms.** McGraw-Hill, 1990.
- DAVI, F. M. **Display technology: Stereo and 3D display technologies.** *Department of Computer Science North Carolina State University Raleigh, NC 27695-8206*, outubro 2003.
- DEROSE, T. D., S. E. J. e H., S. D. **Wavelets for Computer Graphics: A Primer.** TR. 94-09-11, Dep. of Comp. Sci. and Eng., University of Washington, 1994.
- DHOND, U. R. e AGGARWAL, J. K. **Structure from stereo - a review.** *In: IEEE Transactions on Systems, Man and Cybernetics*, 19:1489–1510, 1989.

- DUARTE, M. H. V., SILVA, E. A. B. e M. B. CARVALHO, CARLA L. PAGLIARI, G. V. M. **Multiscale recurrent patterns applied to stereo image coding.** *In: IEEE Transactions on Circuits And Systems for Video Technology. CSVT*, no. 11, 15: 1434–1447, novembro 2005.
- DUDA, R. O. e HART, P. E. ***Pattern Classification and Scene Analysis.*** John Wiley and Sons, 1973.
- ELLINAS, J. N. e SANGRIOTIS, M. S. **Stereo video coding based on quad-tree decomposition of b-p frames by motion and disparity interpolation.** *In: IEE Proceedings - Vision, Image and Signal Processing*, 152:639–647, outubro 2005.
- FAUGERAS, O. ***Three - Dimensional Computer Vision, a Geometric Viewpoint.*** MIT Press, Cambridge (MA), 1993a.
- FAUGERAS, O., HOTZ, B., MATTHIEU, H., VIEVILLE, T., ZHANG, Z., FUA, P., THERON, E., MOLL, L., BERRY, G., VUILLEMIN, J., BERTIN, P. e PROY, C. **Real time correlation-based stereo: Algorithm, implementations and applications - a review.** *In: INRIA Technical Report*, 2013, 1993b.
- FAUGERAS, O. e LUONG, Q. T. ***The Geometry of Multiple Images.*** MIT Press, Cambridge (MA), 2001.
- GONZALEZ, R. C. e WOODS, R. E. ***Digital Imagem Fundamentals.*** 1992.
- HARTLEY, R. e ZISSERMAN, A. ***Multiple View Geometry in Computer Vision.*** Cambridge University Press, Cambridge (UK), 2000.
- HASKELL, B. G., PURI, A. e NETRAVALI, A. ***Digital Video: an Introduction to MPEG-2.*** Chapman and Hall, New York, 1997.
- HENDRIKS, E. A. e MAROSI, G. **Recursive disparity estimation algorithm for real-time stereoscopic video applications.** *In: Proceedings of the 1996 IEEE International Conference on Image Processing, ICIP'96*, págs. 891–894, 1996.
- HHI. *In: Heinrich-Hertz-Institut, Berlin, Germany.*
- HILTON, M. L., J. B. e SENGUPTA, A. ***Compressing Still and Moving Images with Wavelets.*** Multimedia Systems and Eng., vol. 2(3), 1994.
- HONG, B. **Introduction to H.264 multimedia communications laboratory.** *In: University of Texas at Dallas*, novembro 2002.
- HOOF, W. e AHUJA, N. **Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection.** *In: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):121–136, 1989.
- HORN, B. K. ***Robot Vision, MIT Press.*** McGraw-Hill, 1986.
- HORN, B. K. e SCHUNK, B. G. **Determining optical flow.** *In: Artificial Intelligence*, 17:185–204, 1981.

- HOWARD, I. P. e ROGERS, B. J. *Binocular Vision and Stereopsis*. Oxford University Press, New York, 1995.
- ISHIKAWA, H. e GEIGER, D. **Occlusions, discontinuities, and epipolar lines in stereo**. In: *European Conference on Computer Vision*, págs. 232–248, 1998.
- ISO. **Coding of moving pictures and associated audio for digital storage media at up to about 1,5 mbit/s - Part 2: Video, ISO/IEC 11172-2**. In: <http://www.iso.org>, 1993.
- ISO. **MPEG-4 overview - (v.21 jeju version)**. In: *International Organisation for Standardisation, ISO/IEC JTC1/SC29/WG11 N4668, Coding of Moving Pictures and Audio*, março 2002.
- ITU-T. **Video coding for low bit rate communication, ITU-T recommendation H.263, version 1: November 1995; version 2, January 1998; version 3, November 2000**. In: <http://www.itu.int>, novembro 1995.
- ITU-T. **H.264 (advanced video coding for generic audiovisual services), H.264.1 (conformance specification for H.264 advanced video coding) e H.264.2 (reference software for H.264 advanced video coding)**. In: <http://www.itu.int>, março 2005.
- ITU-T e 1, I. J. **Generic coding of moving pictures and associated audio information - Part 2: Video, ITU-T Recommendation H.262 - ISO/IEC 13818-2 (MPEG-2)**. In: <http://www.itu.int> and <http://www.iso.org>, novembro 1994.
- JAIN, A. K. *Fundamentals of Digital Image Processing*. Prentice-Hall International, 1989.
- JENKIN, M. R. M. e JEPSON, A. D. **Recovering local surface structure through local phase difference measurements**. In: *CVGIP: Image Understanding*, 59: 72–93, 1994.
- JVT. **Joint model reference software version 9.2 e 9.4**. *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG*.
- KANADE, T. e OKUTOMI, M. **A stereo matching algorithm with an adaptive window: Theory and experiment**. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):1207–1212, 1994.
- KLETTE, R., SCHLUNS, K. e KOSCHAN, A. *Computer Vision - Three Dimensional Data from Images*. Springer, 1998.
- KOLMOGOROV, V. e ZABIH, R. **Visual correspondence with occlusions using graph cuts**. In: *International Conference on Computer Vision*, págs. 508–515, 2001.
- KOLMOGOROV, V. e ZABIH, R. **Multi-camera scene reconstruction via graph cuts**. In: *7th European Conference on Computer Vision, volume III of LNCS*, 2352: 82–96, maio 2002.

- KONRAD, J. e DUBOIS, E. **Bayesian estimation of motion vector fields.** *In: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):910–927, 1992.
- KOPERNIK, A., SAND, R. e CHOQUET, B. **"The Future of Three-dimensional TV",** *in Signal Processing of HDTV, IV, eds. E. Dubois and L. Chiariglione, 17-29.* Elsevier Science Publishers B. V., 1993.
- LABVISIONE. <http://labvisione.deis.unibo.it/smattoccia/stereo.htm>.
- LIM, J. S. **Two-Dimensional Signal and Image Processing.** Prentice-Hall Signal Processing Series, Alan V. Oppenheim Series Editor, 1990.
- LIN, C., FEI, E. e CHEN, Y. **Rate-distortion constrained quadtree segmentation for stereoscopic video coding.** *In: Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, ISCAS '99.*, 4:311–314, junho 1999.
- LIPTON, L. **Compatibility issues and selection devices for stereoscopic television.** *In: IEEE Image Communication*, 4:15–20, 1991.
- M. D. ADAMS, H. MAN, F. K. e EBRAHIMI, T. **JPEG 2000: The next generation still image compression standard.** *In: ISO/IEC JTC 1/SC 29/WG 1 N 1734*, junho 2000.
- MAIMONE, M. W. **Characterizing stereo matching problems using local spatial frequency.** *In: Ph.D. Thesis, School of Computer Science, Carnegie Mellon University, USA*, 1996.
- MAIMONE, M. W. e SHAFER, S. A. **Modeling foreshortening in stereo vision using local spatial frequency.** *In: Technical Report CMU-CS-95-104, School of Computer Science, Carnegie Mellon University, USA*, 1995.
- MCL. **Multimedia communications laboratory.** *In: Department of Electronic Systems Engineering, University of Essex, U.K.*
- MOTOKI, T., ISONO, H. e YUYAMA, I. **Present status of three-dimensional television reserch.** *In: IEEE Proceedings*, 83:1009–1021, 1995.
- OHTA, Y. e KANADE, T. **Stereo by intra and inter-scanline search using dynamic programming.** *In: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, 1985.
- PAGLIARI, C. L. **Perspective-View Image Matching in the DCT Domain, Tese de Doutorado.** Department of Electronic Systems Engineering University of Essex, junho, 1999.
- PAPADIMITRIOU, D. V. **Shape and motion analysis from stereo for model-based image coding.** *In: Ph.D. Thesis, Department of Eletronic Systems Engineering, University of Essex*, 1995.

- PEI, S. e LAI, C. **Very low bit-rate coding algorithm for stereo video with spatiotemporal HVS model and binary correlation disparity estimator.** *In: IEEE Journal on Selected Areas in Communications*, 1:98–107, janeiro 1998.
- PEREZ, M. M. **Stereo-Based Image Interpretation Using Cooperative Disparity and Segmentation Analysis, Tese de Doutorado.** Department of Electronic Systems Engineering University of Essex, junho, 2003.
- PERKINS, M. G. **Data compression of stereopairs.** *In: IEEE Transactions Community*, 40:684–696, abril 1992.
- RANDRIAMASY, A. e GAGALOWICS, A. **Region based stereo matching oriented image processing.** *In: Proceedings of the 1991 IEEE International Conference on Computer Vision and Pattern Recognition, CVPR'91*, págs. 736–737, 1991.
- REDERT, A., HENDRIKS, E. e BIEMOND, J. **Correspondence estimation in image pairs.** *In: IEEE Signal Processing Magazine*, págs. 29–46, maio 1999.
- RICHARDSON, I. E. G. **H.264/MPEG-4 Part 10 white paper.** *In: www.vcodex.com*, outubro 2002.
- RICHARDSON, I. E. G. **H.264 and MPEG-4 Video Compression (Video Coding for Next-Generation Multimedia).** John Wiley and Sons, The Atrium, Southern Gate, Chichester, West Sussex PO198SQ, England, 2003.
- ROSE, V. **Stereoscopic vision and augmented reality.** *In: Scientific Computing and Automation*, junho 1993.
- ROY, S. **Stereo without epipolar lines: a maximum flow formulation.** *In: International Journal of Computer Vision*, 1(2):1–15, 1999.
- ROY, S. e COX, I. **A maximum-flow formulation of the n-camera stereo correspondence problem.** *In: International Conference on Computer Vision*, 1998.
- SANGER, T. D. **Stereo disparity computation using gabor filters.** *In: Biological Cybernetics*, 59:405–418, 1988.
- SCHAFFER, R., WIEGAND, T. e SCHWARZ, H. **The emerging H.264/AVC standard.** *In: EBU TECHNICAL REVIEW*, janeiro 2003.
- SCHALKOFF, R. J. **Digital Image Processing and Computer Vision.** John Wiley and Sons, 1989.
- SCHARSTEIN, D. e SZELISKI, R. **A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.** *In: International Journal of Computer Vision*, 47(1):7–42, 2002.
- SETHURAMAN, S., G.JORDAN, A. e SIEGEL, M. W. **Multiresolution based hierarchical disparity estimation for stereo image pair compression.** *In: Published in the Proc. of the Symposium on Application of subbands and wavelets*, março 1994.

- SHAWMIN, L. e SHIJUN, S. **System and method for three-dimensional video coding.** In: *US Patent and Trademark Office, Patent Application Full Text and Image Database, United States Patent Application*, abril 2005.
- SULLIVAN, G. J., TOPIWALA, P. e LUTHRA, A. **The H.264/AVC Advanced Video Coding standard: Overview and introduction to the fidelity range extensions.** In: *SPIE Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard: H.264/AVC*, agosto 2004.
- SUN, S. e LEI, S. **Stereo-view video coding using H.264 tools.** In: *Proceedings of SPIE, Image and Video Communications and Processing*, 5685:177–184, março 2005.
- SUN, S., LEI, S. e NOMURA, T. **Stereo video coding support in H.264.** In: *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)*, 2003.
- TOMASI, C. e MANDUCHI, R. **Stereo without search.** In: *Technical Report STAN-CS-TR-95-1543, Stanford University, USA*, 1995.
- TRUCCO, E. e VERRI, A. **Introductory Techniques for 3-D Computer Vision.** Prentice Hall, 1998.
- TSENG, B. e ANASTASSIOU, D. **A multi-viewpoint video coding with MPEG-2 compatibility.** In: *IEEE Transactions on Circuits and Systems for Video Technology*, 6:414–418, 1996.
- TZOVARAS, D., GRAMMALIDIS, N. e STRINTZIS, M. G. **Object-based coding of stereo image sequences using 3D motion/disparity compensation.** In: *IEEE Transactions. CSVT, 1997*, 7:312–327, abril 1997.
- VENKATESWAR, V. e CHELLAPPA, R. **Hierarchical stereo and motion correspondence using feature groupings.** In: *International Journal of Computer Vision*, 15:245–269, 1995.
- WANG, L., GANDHI, R., PANUSOPONE, K., YU, Y. e LUTHRA, A. **Mb-level adaptive frame/field coding for JVT.** In: *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 2nd Meeting: Geneva, CH.*, janeiro 2002a.
- WANG, L., PANUSOPONE, K., GANDHI, R., YU, Y. e LUTHRA, A. **Adaptive frame/field coding for JVT video coding.** In: *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 2nd Meeting: Geneva, CH.*, janeiro 2002b.
- WIEGAND, T., SULLIVAN, G. J., BJONTEGAARD, G. e LUTHRA, A. **Overview of the H.264/AVC video coding standard.** In: *IEEE Transactions on Circuits and Systems for Video Technology*, julho 2003.

- WOJCIECH, M. e PFISTER, H. **3DTV: A scalable system for real-time acquisitions, transmission and autostereoscopic display of dynamic scenes.** *Mitsubishi Electric Research Laboratories, TR2004-067 - <http://www.merl.com>*, dezembro 2004.
- WOODS, A. J., MERRITT, J. O., BENTON, S. A. e BOLAS, M. T. **Stereoscopic displays and virtual reality systems xi.** *In: Proceedings of SPIE, 5291*, janeiro 2004.
- YAN, L., ZHAOYANG, Z. e PING, A. **Stereo video coding based on frame estimation and interpolation.** *In: IEEE Transactions on Broadcasting, 49:14–21*, março 2003.
- ZABIH, R. e WOODFILL, J. **Non-parametric local transforms for computing visual correspondence.** *In: Proceedings of the Third European Conference on Computer Vision*, págs. 150–158, 1994.

## 7 APÊNDICES

## 7.1 APÊNDICE 1: IMAGENS ORIGINAIS

Neste apêndice são mostradas as imagens de teste, pares e seqüências estéreo que foram utilizadas nas simulações desta dissertação.

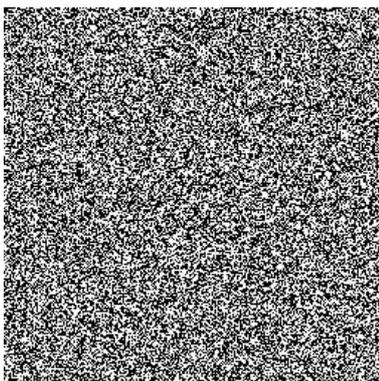
### **Imagem de Teste.**



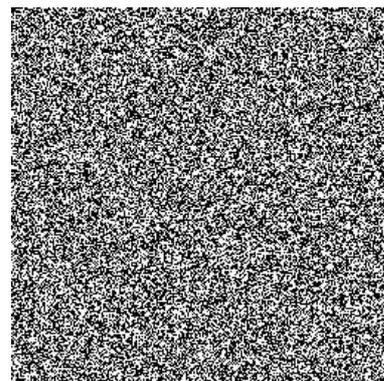
Lena.

FIG. 7.1: Imagem de Teste Lena (512x512).

### **Par Estéreo.**



Quadro 25, vista esquerda - RANDOT.



Quadro 25, vista direita - RANDOT.

FIG. 7.2: Par estéreo - RANDOT (256x256).

## Seqüências Estéreo



(a) Quadro 25, vista esquerda - MAN.



(b) Quadro 25, vista direita - MAN.



(c) Quadro 25, vista esquerda - SAXO.



(d) Quadro 25, vista direita - SAXO.

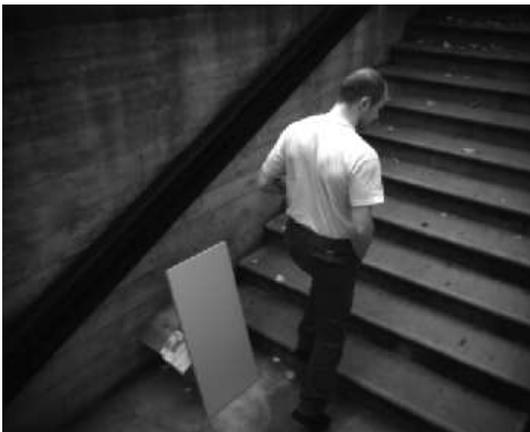
FIG. 7.3: Pares estéreo - MAN (384x384) e SAXO (352x288).



(a) Quadro 90, vista esquerda - INDOOR.



(b) Quadro 90, vista direita - INDOOR.



(c) Quadro 50, vista esquerda - OUTDOOR.



(d) Quadro 50, vista direita - OUTDOOR.

FIG. 7.4: Pares estéreo - INDOOR e OUTDOOR (352x288).

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)